(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
1. März 2001 (01.03.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/15141 A1

(51) Internationale Patentklassifikation[7]: G10L 17/00

(21) Internationales Aktenzeichen: PCT/DE00/02917

(22) Internationales Anmeldedatum:
25. August 2000 (25.08.2000)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
199 40 567.0 26. August 1999 (26.08.1999) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: SIEMENS AKTIENGESELLSCHAFT [DE/DE]; Wittelsbacherplatz 2, 80333 München (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: KUROPATWINSKI, Marcin [PL/DE]; Herzogstr. 40, 46397 Bocholt (DE).

(74) Gemeinsamer Vertreter: SIEMENS AKTIENGESELLSCHAFT; Postfach 22 16 34, 80506 München (DE).

(81) Bestimmungsstaaten *(national)*: AU, BR, CA, CN, CZ, HU, ID, IL, IN, JP, KR, MX, PL, RU, TR, US.

(84) Bestimmungsstaaten *(regional)*: europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht:
— *Mit internationalem Recherchenbericht.*
— *Vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen.*

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

WO 01/15141 A1

(54) Title: METHOD FOR TRAINING A SPEAKER RECOGNITION SYSTEM

(54) Bezeichnung: VERFAHREN ZUM TRAINIEREN EINES SPRECHERERKENNUNGSSYSTEMS

(57) Abstract: The invention relates to a method of recognizing speakers using the parameters of an LPAS encoder or a parametric encoder for modeling the probability distribution for the speaker classes.

(57) Zusammenfassung: Die Erfindung betrifft ein Verfahren zur Sprechererkennung unter Anwendung von Parametern eines LPAS-Kodierers oder eines parametrischen Kodierers zur Modellierung der Wahrscheinlichkeitsverteilung für die Sprecherklassen.

Beschreibung

VERFAHREN ZUM TRAINIEREN EINES SPRECHERERKENNUNGSSYSTEMS

Die Erfindung betrifft ein Verfahren zum Erkennen von Sprechern anhand deren Stimmen.

Die der Erfindung zugrundeliegende Aufgabe besteht darin, ein Verfahren zum Erkennen von Sprechern anhand deren Stimmen anzugeben, das robust, sicher und zuverlässig ist.

Diese Aufgabe wird erfindungsgemäß durch die im Patentanspruch 1 angegebenen Merkmale gelöst.

Im folgenden wird die Erfindung unter Verwendung eines Flußdiagramms näher beschrieben.

1.
Die Erfindung ermöglicht die Erkennung des Sprechers anhand seiner Stimme. Das Problem der Sprechererkennung besteht darin, zwischen verschiedenen Sprechern zu unterscheiden oder die vorgegebene Sprecheridentität zu überprüfen, wobei die einzige Eingangsinformation die Aufzeichnung der Stimme des Sprechers ist.

Außerdem wird eine Methode vorgeschlagen, die das Überlisten des Zugangssystems verhindert, wenn die Stimme und das Schlüsselwort von Dritten aufgenommen wird.

Bei der Speicherung von komplexen Wahrscheinlichkeitsverteilungen für die Sprachparameter eines Sprechers muß zwischen Genauigkeit und Speicherbedarf ein Kompromiss geschlossen werden. Deswegen werden Methoden der Speicherung der Wahrscheinlichkeitsverteilungen vorgeschlagen, die abhängig von der Anzahl der Sprecher einsetzbar sind.

2.
Die Sprechererkennung wurde bisher z.B. mit Hilfe von Hidden-Markov Modellen oder durch Vektorquantisierung gelöst, siehe Literatur [1].

3.
Die Erfindung löst das Problem der Sprechererkennung basierend auf den Parametern einer Analyse durch Synthese Kodierers mit der Linearen Prädiktion (LPAS) [1] (z.B. eines Harmonic Vector Excited Codecs [5] oder Waveform Interpolation Codec [4]). Die bisher verwendeten Parameter des Sprachsignals wie z.B. Cepstrale AR Parameter bringen keine zufriedenstellende Lösung des Problems. Deswegen muß auf andere Parameter zugegriffen werden wie z.B. Parameter der Anregung des Vokaltraktes, die sprecherabhängige und zugleich weitgehend phonemenunabhängige Information tragen.

Darüber hinaus wird die Methode der Schätzung der Wahrscheinlichkeitsverteilung der Kodiererparameter für den jeweiligen Sprecher gegeben, und eine Methode, die das Überlisten des Zugangssystems verhindert.

Sprecheridentifikation

Bei Systemen zur *Sprechererkennung* wird nach den statistischen Prinzipen [2] geprüft, ob der gesprochene Satz von einem der vom Sprechererkennungssystem erfassten Sprecher gesprochen wurde. Dabei gibt es grundsätzlich zwei Arten von Sprechererkennungssystemen, die textabhängigen und die textunabhängigen Systeme. Für die in der Erfindung beschriebene Prozedur wird die Textunabhängigkeit des System durch eine erweiterte Trainingsphase erreicht, in der der Sprecher ein vielfältiges Material aufzeichnen muß und die Wahrscheinlichkeitsverteilungen der erwähnten Sprachsignalparameter aus dem gesamten Sprachmaterial bestimmt. Das Trainieren eines textabhängigen Systems ist eine einfachere Aufgabe, weil das Sprachmaterial, das vom Sprecher während der Nutzungsphase gesprochen wird, auf einige Schlüsselworte oder bestimmte

Sätze begrenzt ist. Die Vorbereitungsphase wird so lange durchgeführt, bis das System sicher die Stimme des Sprechers erkennt.

Die Aufgabe der *Sprecheridentifikation* ist in Bild 2 dargestellt.

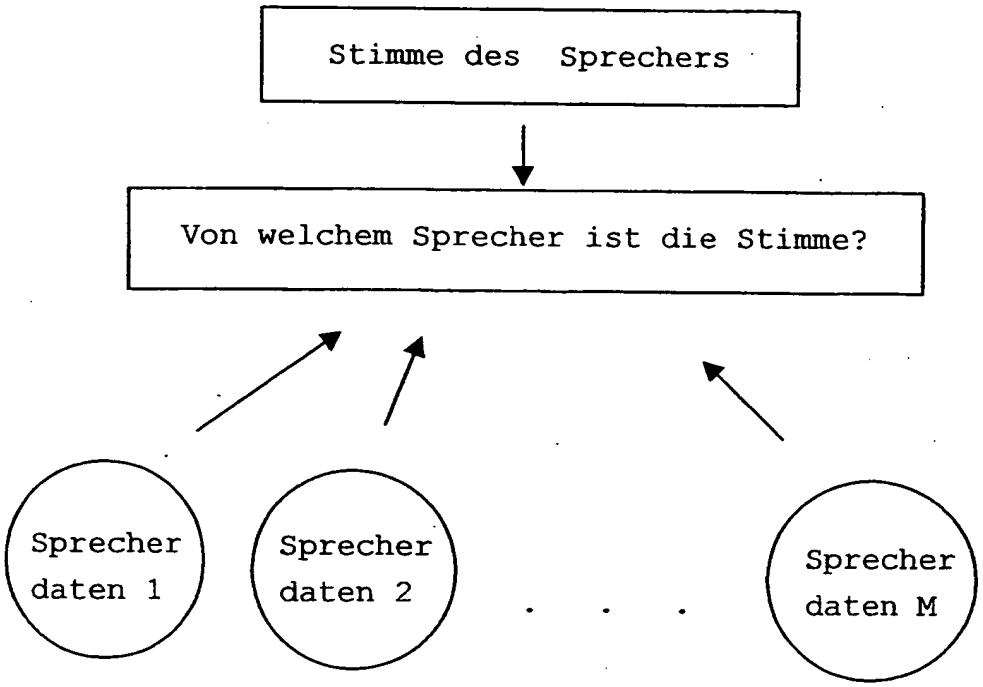


Bild 2. Problem der Sprecheridentifikation

Die *Sprecheridentifikation* wird als ein Problem der Multiplen Detektion behandelt [2]. Die zu unterscheidenden Klassen, eine für jeden Sprecher, das vom System erkannt werden soll, werden als sp_i i = 1..M bezeichnet, mit M - Anzahl der von dem Sprechererkennungssystem erfassten Sprecher. Die Sprechererkennung basiert auf den aufgezeichneten Sprachsignalen der jeweiligen Sprecher. Das Sprachsignal wird segmentiert in die Signalrahmen $\mathbf{x} = [x(1)..x(K)]$ (z.B. für einen Signalrahmen von 20 ms Länge und eine Abtastfrequenz von 8 kHz beträgt K = 160). Die Segmentierung liefert die Sprachsignalrahmen $\mathbf{x}(1)..\mathbf{x}(N)$, wobei N von der Gesamtlänge des von dem Sprecher gesprochenen Satzes oder Schlüsselwortes abhängt. Die Entscheidung über den Sprecher wird aus den Wahrscheinlichkeiten oder Wahrscheinlichkeitsdichten (zusammen als Wahrscheinlichkeitsscores bezeichneten) getroffen, daß die Vektoren der Abtastwerte $\mathbf{x}(l)$ $l = 1..N$ der Klasse sp_i zugehören. Das statistisch optimale Entscheidungsschema wählt die Klasse sp_i mit dem höchsten Wahrscheinlichkeitswert bei gegebenen $\mathbf{x}(l)$, $l = 1..N$. D.h. der Vektor $\mathbf{x}(l)$ wird der Klasse sp_j zugeordnet, für die:

$$p(\mathbf{x}(1)...\mathbf{x}(N) \mid sp_j) > p(\mathbf{x}(1)...\mathbf{x}(N) \mid sp_i) \quad \text{für alle } j \neq i$$

Sprecherverifikation

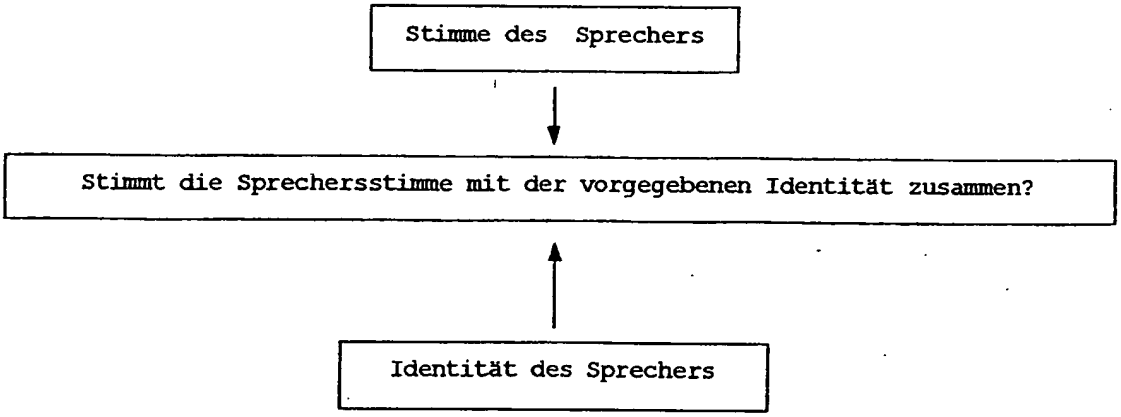


Bild 3. Problem der Sprecherverifikation

Problem der *Sprecherverifikation* besteht darin, die vorgegebene Identität des Sprechers anhand seiner Stimme zu überprüfen. Dies entspricht der auf dem Bild 3. abgebildeten Situation.

Der Prozeß der Sprecherverifikation verläuft auf ähnliche Weise wie der bei der Sprecheridentifikation, d.h. es wird ebenfalls die Segmentierung des gesprochenen Satzes durchgeführt. Danach wird jedoch keine Klassifizierung der Stimme gemacht, sondern für die vorgegebene Sprecheridentität ein Wahrscheinlichkeitsscore berechnet und mit einer Schwelle verglichen. Die Identität des Sprechers wird also anhand seiner Stimme bestätigt, wenn:

$$p(x(1)..x(N) \mid sp_j) > schwelle$$

wobei sp_j der vorgegebenen Sprecheridentität entspricht. Die Schwelle muß entsprechend hoch gesetzt werden, um die Situation zu vermeiden, in der ein Sprecher mit einer anderer Identität als die vorgegebene zugelassen/autorisiert wird.

LPAS Kodierer

Die heute eingesetzten Sprachkodierverfahren basieren vorwiegend auf dem Analyse-durch-Synthese Verfahren mit einem LPC-Synthesefilter [2]. Die Sprachkodierung wird in diesen Verfahren durch Wiederholung der Kodierungs- und Dekodierungs-Operationen solange optimiert, bis der optimale Parametersatz für den gegebenen Sprachabschnitt gefunden wird.

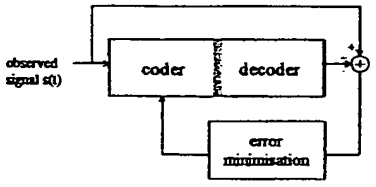


Bild 4: Schema eines LPAS Kodierers

Einer der am meisten verwendeten Typen des LPAS Kodierers ist der CELP Kodierer. Eine relativ neue Entwicklung ist der Harmonic Vector Excited Codec mit einer besonders für die beschriebene Aufgabe geeigneter Form der Anregungssignale. Synthesemodell eines CELP Kodierers ist in Bild 4 dargestellt. Das Synthesemodell definiert die Methode der Berechnung des synthetisierten Sprachsignals aus den quantisierten Parametern des Sprachsignals. Im allgemeinen besitzt jeder LPAS Kodierer Parametergruppen:

- Kurzzeitprädiktorparameter. Die Kurzzeitprädiktorparameter werden in der Regel mit Hilfe klassischer LPC Analyse berechnet, wobei die Korrelations-Methode oder die Kovarianz-Methode der Linearen Prädiktion angewendet wird [3]. Für Signalrahmen der Länge von 20 bis 30 ms und eine Abtastrate von 8kHz werden 8-10 LPC Koeffizienten verwendet. Die Kurzzeitprädiktorparameter können in verschiedenen Formen (z.B. die Reflexionskoeffizienten oder als Line Spectrum Frequencies LSF) auftreten, abhängig davon, welche Darstellung sich besser quantisieren läßt. Es hat sich gezeigt, daß die LSF Koeffizienten am besten zur Quantisierung geeignet sind und diese Form der Prädiktionskoeffizienten wird in der Regel verwendet. Die Kuzrzeitprädiktorparameter werden in einer open-loop Prozedur berechnet, d.h. ohne der auf dem Bild 1 dargestellten gesamten Optimierung mit den anderen Parametern bezüglich des Synthesefehlers.

- Langzeitprädiktorparameter. Langzeitprädiktorparameter werden in einem Filter verwendet, der die Grundfrequenz des Sprachsignals synthetisiert. Es wird am meisten ein Langzeitprädiktor mit einem Filterkoeffizient und einem Parameter für die Grundperiode des Sprachsignals. Ein Langzeitprädiktor mit den Parametern $\mathbf{b} = [b, N]$ ist ein Teil der Abb. 2. Die Langzeitprädiktorparameter werden ebenfalls in einer open-loop Prozedur berechnet ohne eine Gesamtoptimierung mit den anderen Parametern. In manchen Ko-

dierern wird manchmal eine verfeinerte Suche nach den Langzeitprädiktorparametern in einer closed-loop Prozedur durchgeführt.

- Die Parameter der Anregung. In einem CELP Kodierer werden die 5-10ms Subrahmen des Restsignals in einer closed-loop Prozedur vektorquantisiert. Die gesendeten Parameter ermöglichen auf der Dekoderseite die Wiederherstellung der Signalformen aus dem gespeicherten Codebuch.

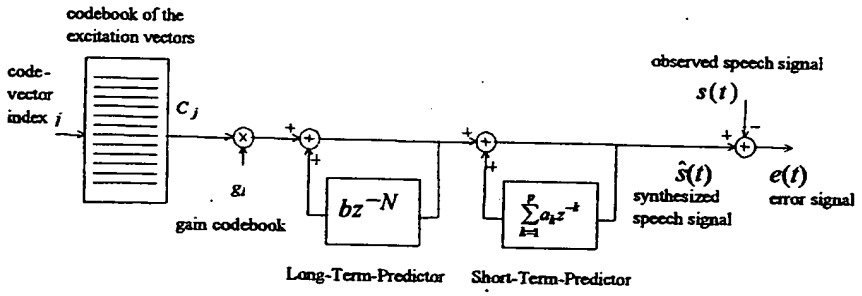


Bild 5.: Synthese-Modell eines CELP Kodierers

In einem HVXC Codecs wird der Ausgang aus dem LPC Analyse Filter in die Frequenzdomäne transformiert und die grundperiodennormalisierte Spektraleinhüllende vektorquantisiert.

<u>Sprechererkennung mit den Parametern eines LPAS Kodierers</u>

Die Parameter eines Sprachkodierers beschreiben ausführlich die möglichen Sprachsignale mit einer wesentlich reduzierten Anzahl der Parameter im Vergleich zur Darstellung des Sprachsignals als eine Sequenz der Abtastwerte.

Die Dekomposition des Sprachsignals in die erwähnten Parametergruppen kann auf verschiedene Weise zur Sprechererkennung verwendet werden. Die Methoden zur Berechnung der Parameter und Synthese des Sprachsignals implizieren die Methoden der Schätzung der Wahrscheinlichkeitsdichten (bzw. der Wahrscheinlichkeiten für die Parameter, die als diskrete Wahrscheinlichkeitsvariablen betrachtet werden). Die in einer closed-loop Prozedur bestimmt werden, sollen eigentlich als

diskrete Wahrscheinlichkeitsvariablen betrachtet werden, weil es nicht möglich ist, für solche Parameter die Volumen der Parameterraumesregionen des Vektorquantizierers zu verbinden. Dies betrifft insbesondere die Anregungsparameter. Die Schätzung der Wahrscheinlichkeitsverteilungen für solche Parameter wird durch die Berechnung von relativen Häufigkeiten der Parameter/Codevektoren im Trainingsatz bestimmt.

Die in einer open-loop Prozedur im Kodierer berechnet werden, sind zuerst in einer nichtquantisierten Form verfügbar und dann erst quantisiert, wobei in der Regel die Vektorquantisierung verwendet wird. Für solche Parameter können die Wahrscheinlichkeitsdichten aus dem Trainingssatz geschätzt werden. Dieser Ansatz wird vor allem für die Kurzzeitprädiktorparameter angewendet.

Die Schätzung der Wahrscheinlichkeitsdichten basiert auf der Histogramm Methode [6]. Diese Methode benötigt die Kenntnisse der Volumen der mit den quantisierten Punkten verbundenen Regionen des Parameterraumes.

Eine Methode der Speicherung von Wahrscheinlichkeitverteilungen ergibt sich, wenn die möglichen Codevektoren für die Sprachsignalparameter einmal für die ganze Population gespeichert werden, was dem Fall entspricht, daß die Quantisierungsstufen/Codevektoren aus der Datenbank bestimmt, die die Aufzeichnungen von vielen Sprechern beinhaltet, einmal bestimmt werden. Die Wahrscheinlichkeitsverteilungen der Parameter für die Sprecher werden dann zusammen mit den Indizien der Codevektoren für die Parameter im System gespeichert. Sie eignet sich für große Systeme mit sehr vielen Anwendern (ATM, Zugangssysteme in Betrieben).

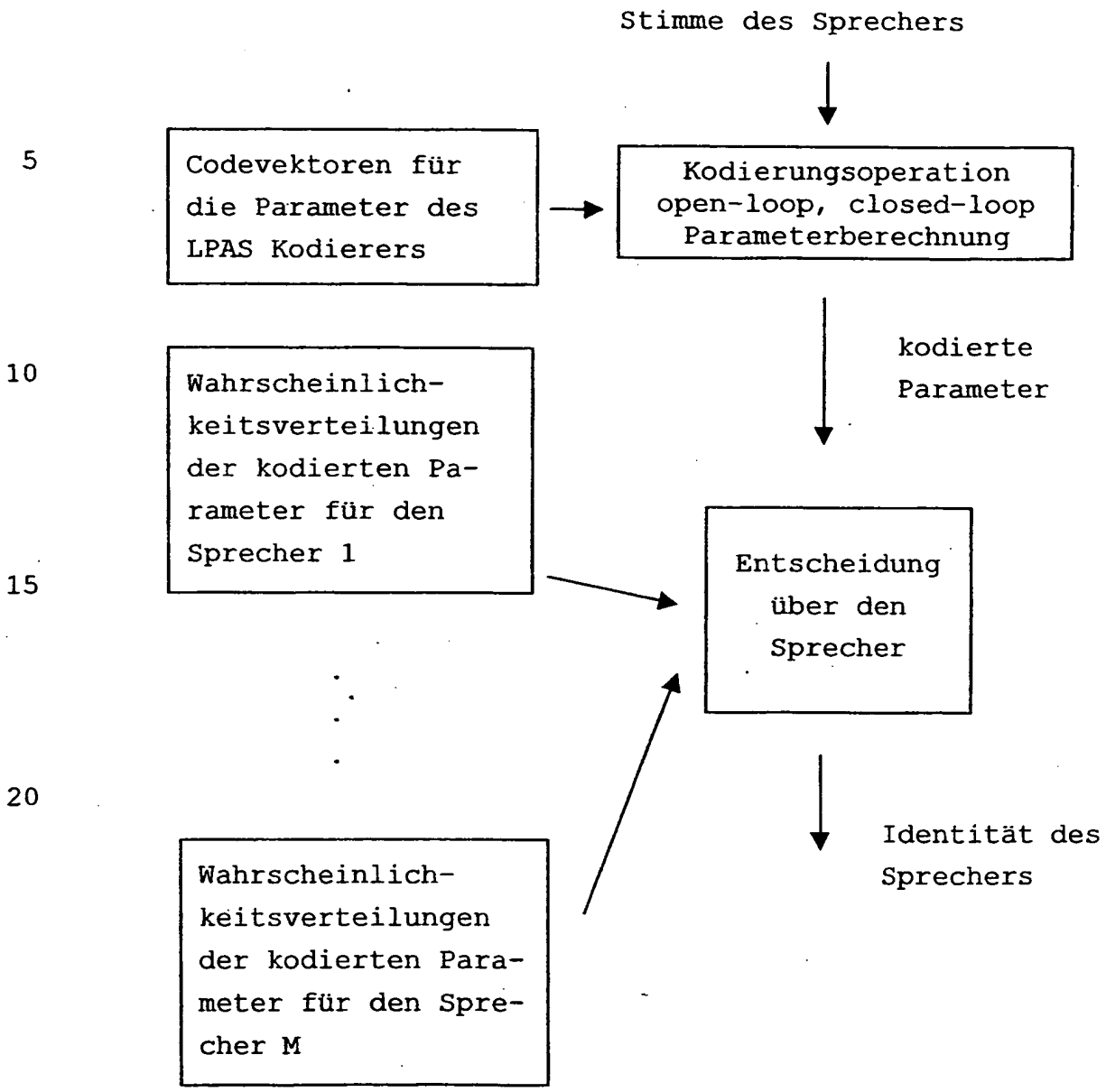


Bild 6. Sprecheridentifikation mit den Parameter eines LPAS Kodierers

Eine andere Methode ergibt sich, wenn die Codevektoren für die Parameter für jeden Sprecher einzeln trainiert werden. Die Codevektoren werden dann zusammen mit den Werten der Wahrscheinlichkeitsdichten an den durch die Codevektoren bestimmten Punkten des Parameterraumes gespeichert. Ein Schema dieser Methode ist auf dem Bild. 7 gezeigt. Diese Methode ist für eine kleine Anzahl von Sprechern bestimmt (z.B. für eine mit der Stimme gesteuerte Tür in der Wohnung)

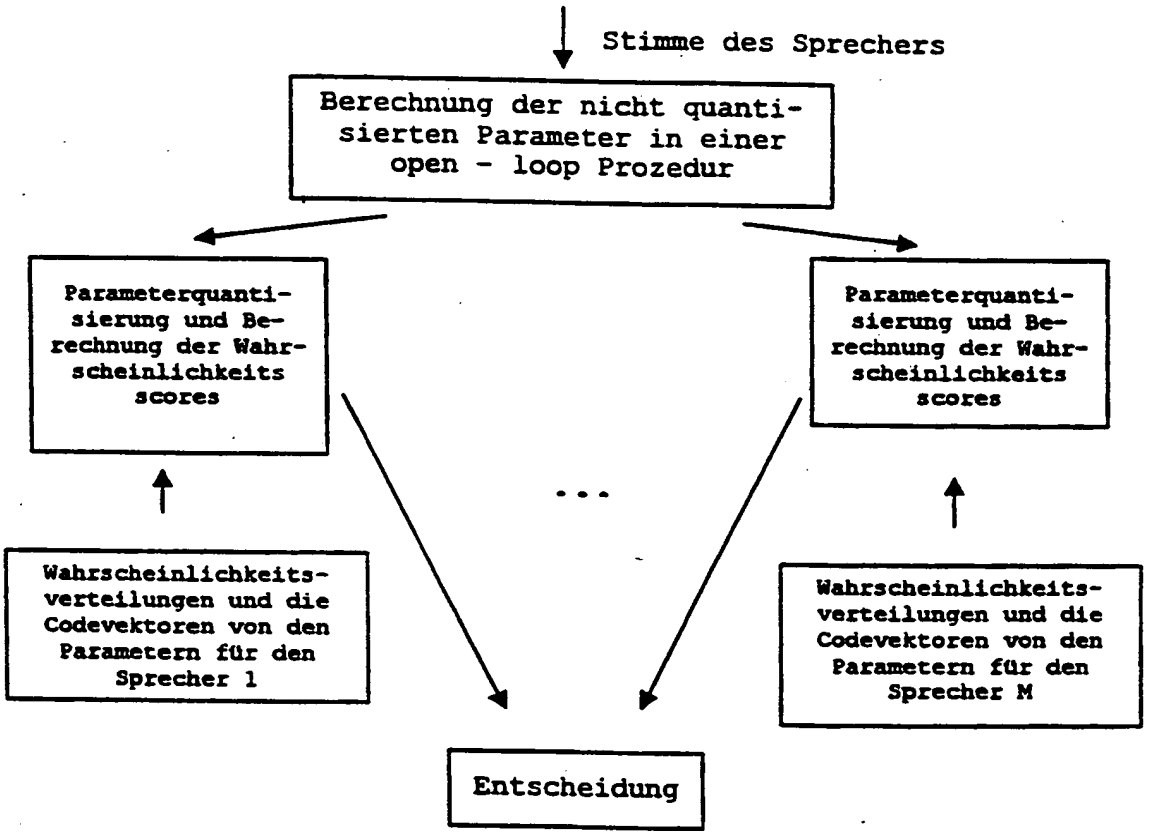


Bild 7. Sprecheridentifikation mit den Parametern ein s LPAS Kodierers Wahrscheinlichkeitsdichten werden zusammen mit den Codevekt ren für die Parameter gespeichert

↓ Identität des Sprechers

Trainingsphase eines Sprechererkennungssystems

Die Wahrscheinlichkeitsdichteverteilungen für die Sprecherklassen werden aus dem Trainingsmaterial geschätzt. Für die textabhängige Sprechererkennung (Sprecheridentifikation/Sprecherverifikation) wird ein bestimmter Satz oder Schlüsselwort während der Trainingphase so lange wiederholt bis die Sprechererkennung sicher funktioniert.

Für die textunabhängige Sprecherverifikation muß ein phonetisch ausgewogenes Sprachmaterial aufgenommen werden. Auch in diesem Fall muß die Trainingphase solange wiederholt werden bis die Sprecheridentifikation/verifikation sicher funktioniert.

Das während der Trainingphase aufgenommene Material wird zum Training mehrmals jeweils phasenverschoben verwendet, um das Sprechererkennungssystem unabhängig von der Anfangsphase der aufgezeichneten Stimmen zu machen. Die zum Training verwendeten Daten wird als Trainingsatz TS_{sp_i} bezeichnet, wobei sp_i den Sprecher symbolisiert.

Schätzung der Wahrscheinlichkeitsdichten

Um die erfindungsgemäße Methode zur Schätzung der Wahrscheinlichkeitsdichten der Parameter für die Sprecherklassen zu beschreiben, werden zuerst notwendige Definitionen eingeführt. Die eingeführte Abstraktion des Kodierungsprozesses hat den Vorteil, daß die Schätzung der Wahrscheinlichkeitsdichten auf einfache Weise beschrieben werden kann, ohne auf die sehr komplizierten Operationen im Sprachkodierer in Details einzugehen. Eine detaillierte Beschreibung der Parameterberechung kann in [4] und [5] gefunden werden.

Ein Sprachkodierer arbeitet in Auswerteintervallen. Für jeden Signalrahmen werden in dem Sprachkodierer die im Abschnitt über LPAS Kodierer beschriebene Operationen durchgeführt, die

die Parameter des Sprachsignals für den jeweiligen Rahmen liefern.

Berechnung eines nicht quantisierten Parametervektors $\mathbf{p}$ aus dem Signalrahmen $\mathbf{x}$ in einer open-loop Oprimierungprozedur wird als $\mathbf{p} = K_p(\mathbf{x})$ geschrieben. Die Quantisierung des Parameters wird als: $\hat{\mathbf{p}} = Q_p(\mathbf{p})$ bezeichnet. Die Region im Parameterraum der Parameter $\mathbf{p}$, der im Kodierungsprozess auf den Codevektor $\hat{\mathbf{p}}$ abgebildet wird, wird als $S_{\hat{\mathbf{p}}} = \{\mathbf{p} : Q_p(\mathbf{p}) = \hat{\mathbf{p}}\}$ bezeichnet. Das Volumen von dieser Region wird als $V(S_{\hat{\mathbf{p}}})$ bezeichnet.

Der Satz möglicher Codevektoren für den Parameter $\mathbf{p}$ wird als $C_p = \{\hat{\mathbf{p}}_i; i = 1..N_p\}$ geschrieben mit N_p Anzahl von Codevektoren. Der Satz von Regionen, die mit den Codevektoren verbunden sind, wird als $R_p = \{S_i; i = 1..N_p\}$ bezeichnet. Die Zugehörigkeitsfunktion einer Region S_i wird als:

$$1_{S_i}(\mathbf{p}) = \begin{cases} 1 \text{ für } \mathbf{p} \in S_i \\ 0 \text{ für } \mathbf{p} \notin S_i \end{cases}$$

bezeichnet.

Die Häufigkeit des Vorkommens eines Parameters im Trainingssatz wird mit

$$f_{S_i} = \frac{\text{Anzahl von Parameterwerten aus dem Training Satz } TS_{sp_i} \text{ die in die Region } S_i \text{ fallen}}{\text{Anzahl von Parameterwerten aus dem Training Satz } TS_{sp_i}}$$

berechnet.

Die geschätzte Wahrscheinlichkeitsdichteverteilung wird dann zu:

$$p(\mathbf{p} \mid sp_i) = \sum_{k=1}^{N_p} 1_{S_k}(\mathbf{p}) \frac{f_{S_i}}{V(S_i)}$$

Schätzung der Wahrscheinlichkeiten

Für die Parameter, die als eine diskrete Wahrscheinlichkeitsvariable betrachtet werden, d.h vor allem die Anregung aus dem Codebuch, die in einer closed-loop Prozedur optimiert wird und die Grundperiode des Sprachsignals, werden die Wahrscheinlichkeitsfunktionen (probability mass functions) geschätzt. Diese werden als die Häufigkeiten der gegebenen Pa-

rametercode im Trainingssatz für den jeweiligen Sprecher bestimmt.

Speichern der Wahrscheinlichkeitsverteilungen

Die Sprachparameter in einem Sprachkodierer werden nicht alle gleichzeitig sondern nacheinander berechnet. Es werden z.B. zuerst die Kurzzeitprädiktorparameter berechnet und dann für bereits bekannte Kurzzeitprädiktorparameter die restlichen Parameter bezüglich der Synthese oder des Prädiktionsfehlers optimiert. Dies ermöglicht effektives Speichern der Wahrscheinlichkeitsverteilungen als bedingte Wahrscheinlichkeiten der Codevektoren in einer Baumstruktur. Dies ist möglich dank folgender Abhängigkeit:

$$p(\mathbf{p}_K,\mathbf{p}_L,\mathbf{p}_A \mid sp_i) = p(\mathbf{p}_K \mid sp_i)p(\mathbf{p}_L \mid sp_i,\mathbf{p}_K)p(\mathbf{p}_A \mid sp_i,\mathbf{p}_K,\mathbf{p}_L)$$

$\mathbf{p}_K$ - Vektor von Kurzzeitparameter
$\mathbf{p}_L$ - Vektor von Langzeitparameter
$\mathbf{p}_A$ - Vektor von Anregungsparameter

Eine wesentliche Vereinfachung ergibt sich, wenn die Sprachparameter innerhalb eines Signalrahmens als statistisch unabhängig angenommen werden können. Die obige Formel wird dann zu:

$$p(\mathbf{p}_K,\mathbf{p}_L,\mathbf{p}_A \mid sp_i) = p(\mathbf{p}_K \mid sp_i)p(\mathbf{p}_L \mid sp_i)p(\mathbf{p}_A \mid sp)$$

Die Wahrscheinlichkeitsdichten müssen im System an sehr vielen Punkten im Parameterraum gespeichert werden. Die zum Speichern von Wahrscheinlichkeitsdichten verwendete Bitanzahl ist für die Komplexität des Gesamtsystems kritisch. Für die Wahrscheinlichkeitswerte wird deswegen ein Vektorquantisierer verwendet. Dies ermöglicht die Reduzierung der zum Speichern der Wahrscheinlichkeitsverteilungen verwendeten Bitanzahl.

Systemsicherheit

Um die Überlistung des Systems zu verhindern, wird gleichzeitig mit der Aufzeichnung der Stimme des Sprechers ein Rauschen ausgestrahlt, das dem System bekannt ist und aus dem das digitalisierte Sprachsignal subtrahiert wird.

5.

Die Erfindung kann für Anwendungen der Zutrittskontrolle, wie z.B. die mit der Stimme gesteuerte Tür, oder als Verifikation, beispielsweise für Bankzugangssysteme genutzt werden. Die Prozedur kann als ein Programmodul auf einem Prozessor implementiert werden, der die Aufgabe der Sprechererkennung im System realisiert.

[1] S.Furui, „Recent advances in speaker recognition", Pattern Recognition Letters, Tokyo Inst. of Technol., 1997

[2] P.Vary, U.Heute, W.Hess, *Digitale Sprachsignalverarbeitung*, B.G.Teubner Stuttgart, 1998

[3] K.Kroschel, *Statistische Nachrichtentheorie*, 3rd ed., Springer-Verlag, 1997

[4] W.B.Kleijn, K.K.Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995

[5] ISO/IEC 14496-3, MPGA-3 HVXC Speech Coder description

[6] Prakasa Rao, Functional Estimation, Academic Press, 1982

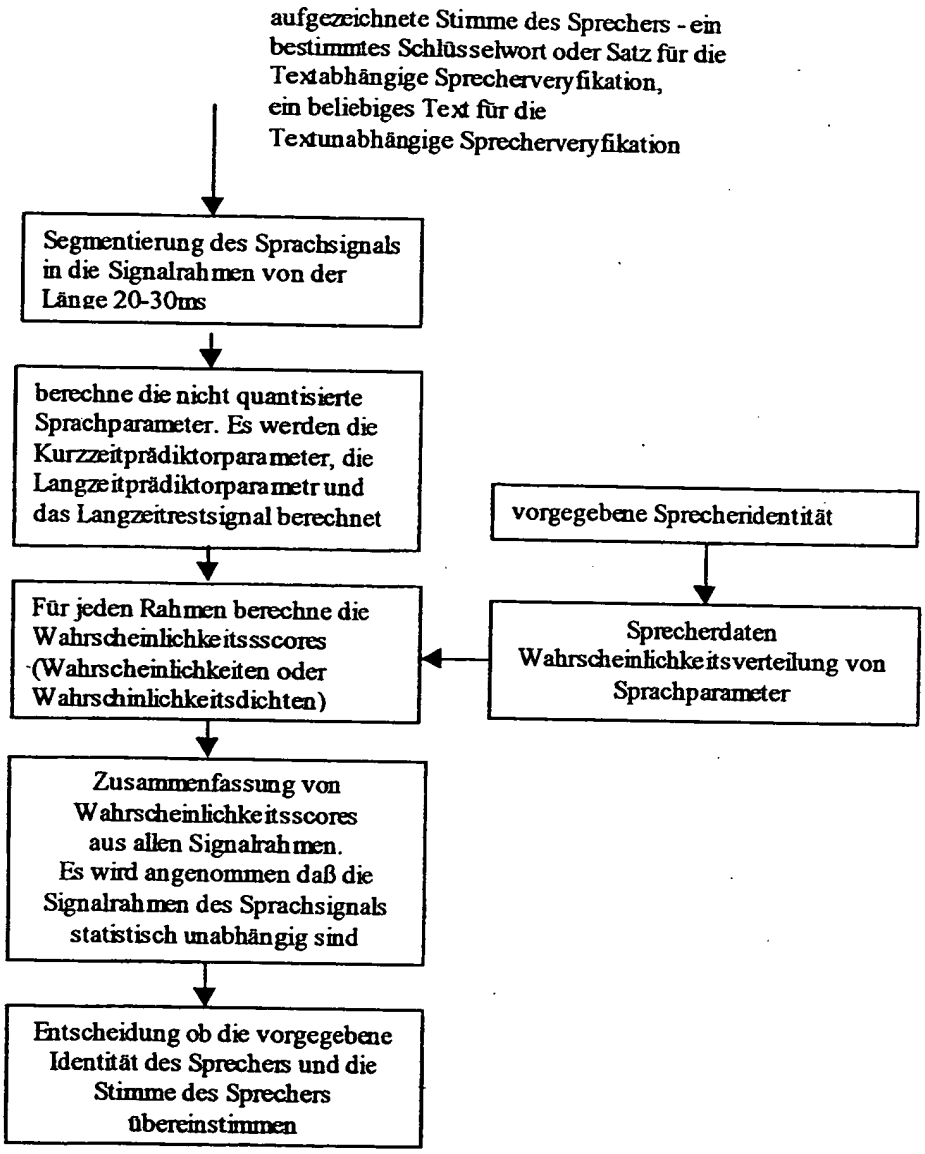


Fig . 1 Sprecherveryfikation mit Verwendung von den Parameter eine LPAS Kodierer

Patentansprüche

1. Verfahren zum Erkennen von Sprechern anhand deren Stimmen mit folgenden Merkmalen:

(a) in einer Vorbereitungsphase,

(a1) werden von M Sprechern jeweils k textabhängige oder textunabhängige Referenzsprachäußerungen, die einen sprecherbezogenen Trainingssatz bilden, in erste Sprachsignalrahmen der Länge L segmentiert,

(a2) werden die ersten Sprachsignalrahmen einem auf linearer Prädiktion basierenden Analyse-durch-Synthese-Kodierer zugeführt,

(a3) wird in dem Analyse-durch-Synthese-Kodierer für jeden der M Sprecher und jeweils jeden ersten Sprachsignalrahmen ein erster Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter des Kodierers berechnet, wobei die Parameter dann ein sprecherbezogenes Trainingsmaterial bilden,

(a4) wird in dem Analyse-durch-Synthese-Kodierer für jeden der M Sprecher und jeweils jeden ersten Sprachsignalrahmen die Häufigkeit des jeweiligen Vorkommens des ersten Kurzzeitprädiktorparameters, Langzeitprädiktorparameters und/oder Anregungsparameters des Kodierers in dem sprecherbezogenen Trainingssatz bzw. die Wahrscheinlichkeitsdichten, mit der der erste Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter in dem sprecherbezogenen Trainingssatz enthalten ist, berechnet,

(a5) werden die berechneten Häufigkeiten bzw. Wahrscheinlichkeitsdichten sprecherbezogen als Sprecherdaten gespeichert,

(b) in einer simulierten Nutzungsphase der Trainingsphase,

(b1) wird eine textabhängige oder textunabhängige Simulationssprachäußerung eines m-ten Sprechers mit m=1..M in zweite Sprachsignalrahmen der Länge L segmentiert,

(b2) werden die zweiten Sprachsignalrahmen dem Analyse-durch-Synthese-Kodierer zugeführt,

(b3) wird in dem Analyse-durch-Synthese-Kodierer für den m-ten Sprecher und jeweils jeden zweiten Sprachsignalrahmen ein

zweiter Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter des Kodierers berechnet,
(b4) werden für jeden zweiten Sprachsignalrahmen aus dem berechneten zweiten Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter und den für den m-ten Sprecher in der Vorbereitungsphase gespeicherten Sprecherdaten erste Wahrscheinlichkeitstreffer berechnet, die angeben, mit welcher Wahrscheinlichkeit der zweite Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter mit dem ersten Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter übereinstimmt,
(b5) werden die ersten Wahrscheinlichkeitsscores aus allen zweiten Sprachsignalrahmen zusammengefaßt,
(b6) wird überprüft, ob die zusammengefaßten ersten Wahrscheinlichkeitsscores größer einer vorgegebenen ersten Schwelle sind, die Stimme des m-ten Sprechers bestätigt, wenn die zusammengefaßten ersten Wahrscheinlichkeitsscores größer als die vorgegebene erste Schwelle sind oder die Vorbereitungsphase solange für weitere i Referenzsprachäußerungen des m-ten Sprechers durchgeführt, bis die Stimme des m-ten Sprechers bestätigt wird, wenn die zusammengefaßten ersten Wahrscheinlichkeitsscores kleiner gleich oder kleiner der vorgegebenen ersten Schwelle sind,
(c) in einer Nutzungsphase
(c1) wird eine textabhängige oder textunabhängige Nutzsprachäußerung des m-ten Sprechers mit m=1.. M in dritte Sprachsignalrahmen der Länge L segmentiert,
(c2) werden die dritten Sprachsignalrahmen dem Analyse-durch-Synthese-Kodierer zugeführt,
(c3) wird in dem Analyse-durch-Synthese-Kodierer für den m-ten Sprecher und jeweils jeden dritten Sprachsignalrahmen ein dritter Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter des Kodierers berechnet,
(c4) werden für jeden dritten Sprachsignalrahmen aus dem berechneten dritten Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter und den für den m-

ten Sprecher in der Vorbereitungsphase gespeicherten Sprecherdaten zweite Wahrscheinlichkeitstreffer berechnet, die angeben, mit welcher Wahrscheinlichkeit der dritte Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter von dem m-ten Sprecher ausgesprochen wurde,

(c5) werden die zweiten Wahrscheinlichkeitstreffer aus allen dritten Sprachsignalrahmen zusammengefaßt,

(c6) wird überprüft, ob die zusammengefaßten zweiten Wahrscheinlichkeitsscores größer einer vorgegebenen zweiten Schwelle sind, die Stimme des m-ten Sprechers wird erkannt, wenn die zusammengefaßten zweiten Wahrscheinlichkeitstreffer größer der vorgegebenen zweiten Schwelle sind oder die Stimme des m-ten Sprechers wird nicht erkannt, wenn die zusammengefaßten zweiten Wahrscheinlichkeitsscores kleiner gleich oder kleiner der vorgegebenen zweiten Schwelle sind.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß
als ein parametrischer Kodierer, insbesondere ein "Harmonic Vector Excited Predictive"-Kodierer oder ein "Waveform Interpolating"-Kodierer verwendet wird.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß
als Analyse-durch-Synthese-Kodierer ein auf linearer Prädiktion basierender Kodierer, insbesondere ein LPAS-Kodierer benutzt wird.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß
die Häufigkeiten bzw. Wahrscheinlichkeitsdichten mit einem Vektorquantisierer mit einer bestimmten, wesentlich reduzierten Bitanzahl quantisiert werden.

5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß

mit der Eingabe der Sprachäußerung des Sprechers in das Sprechererkennungssystem ein dem Sprechererkennungssystem bekanntes Rauschen mit eingegeben wird.

6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß
das miteingegebene Rauschen intern vor der Segmentierung von der Aufnahme der Sprecherstimme subtrahiert wird.

Vorbereitungsgphase des Sprechererkennugnssystems*
(Verlauf für den Sprecher j)

Training eines Textunabhängigen Systems

Aufnahme eines vielfaeltigen phonetisch ausgewogenen Materials von dem j-ten, j=1..M Systemanwender. Eine relativ grosse Anzahl 1..K der Referenzsprachäußerungen

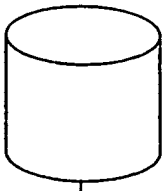
Training eines textabhaengigen Systems

Bestimmte Wortsequenz, ein Satz oder Schluesselwort. Entsprechende Anzahl 1..K der Referenzsprachäußerungen dem j-ten, j=1..M Systemanwender.

Segmentierung des Trainingmaterials in die Signalrahmen $\mathbf{x}(1)...\mathbf{x}(N)$ mit N Abhaengig von der Gesamtlaenge der Sprachäußerungen. $\mathbf{x}(i)=[x(1)...x(L)]$ mit L - Länge des Signalrahmens.

1

Grosse Anzahl von Sprechern >10



Sprachdatenbank Mehrere Stunden an Aufnahmen von verschiedenen Sprecher

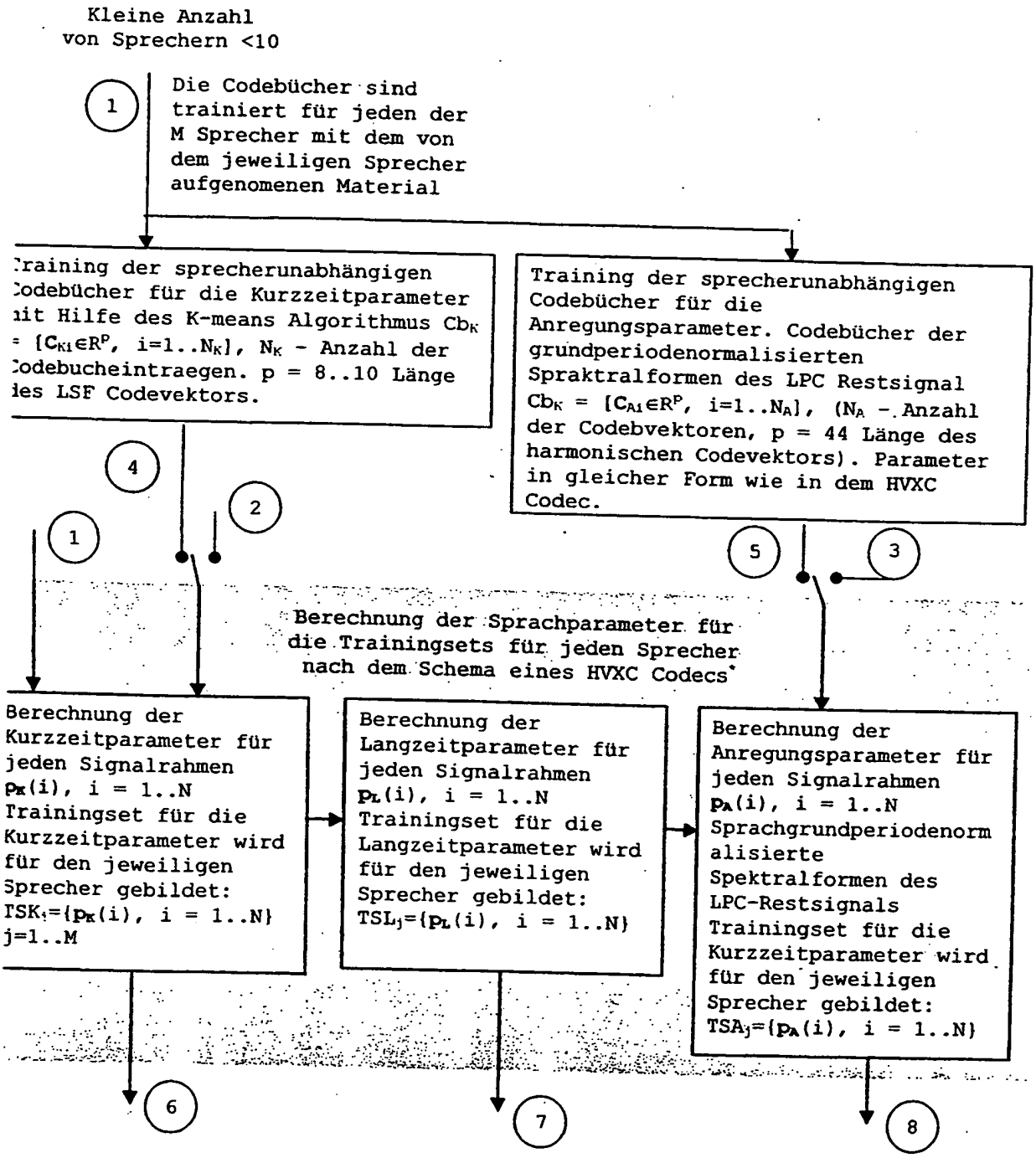
Training der sprecherunabhängigen Codebücher für die Kurzzeitparameter mit Hilfe des K-means Algorithmus $Cb_K = [C_{Ki} \in R^p, i=1..L_K]$, L_K - Anzahl der Codebucheintraegen. p = 8..10 Länge des LSF Codevektors.

2

Training der sprecherunabhängigen Codebücher für die Anregungsparameter. Codebücher der grundperiodenormalisierten Spektralformen des LPC Restsignal $Cb_K = [C_{Ai} \in R^p, i=1..L_A]$, ($L_A$ - Anzahl der Codevektoren, p = 44 Länge des Codevektors). Parameter in gleicher Form wie in dem HVXC Codec.

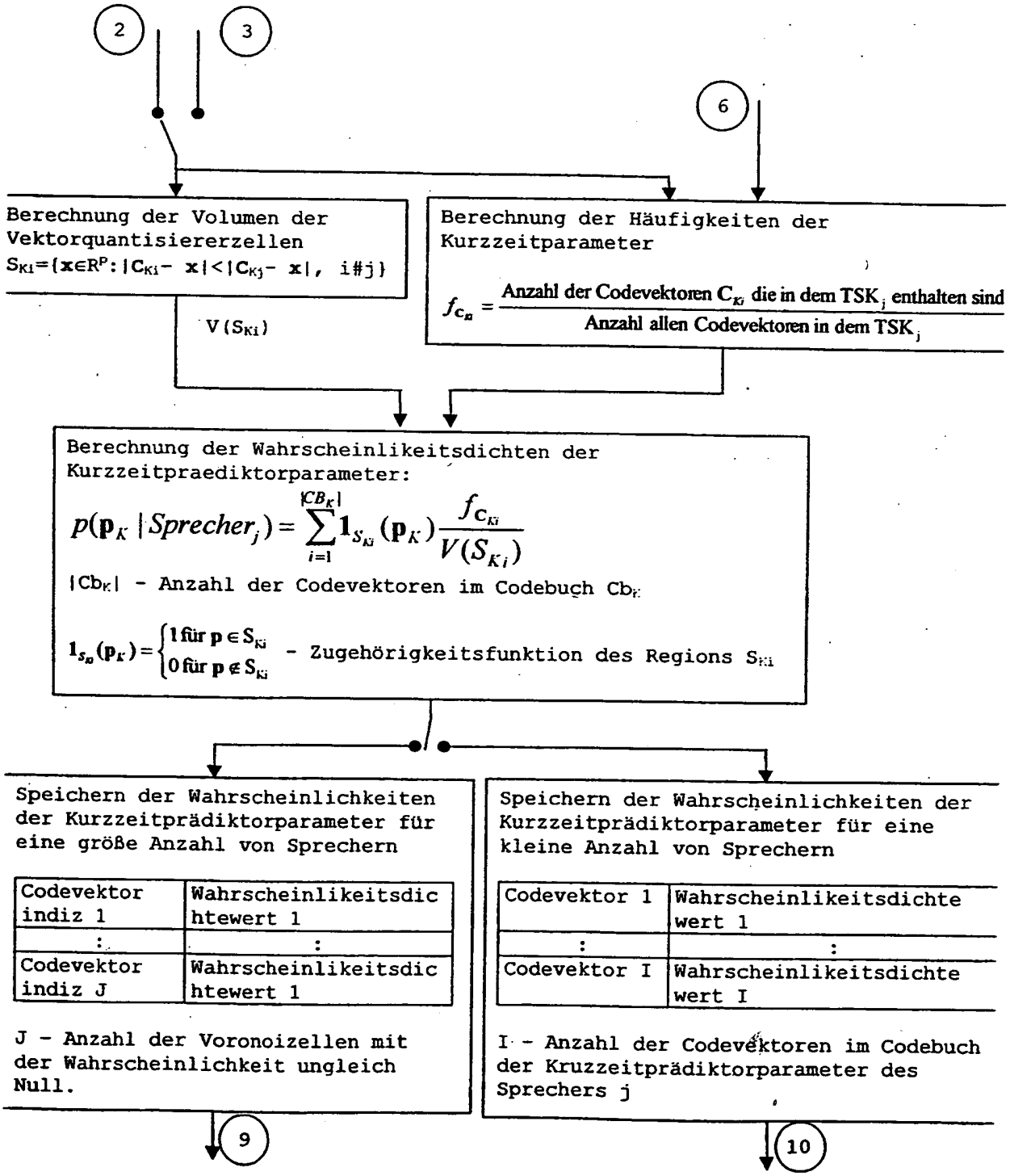
3

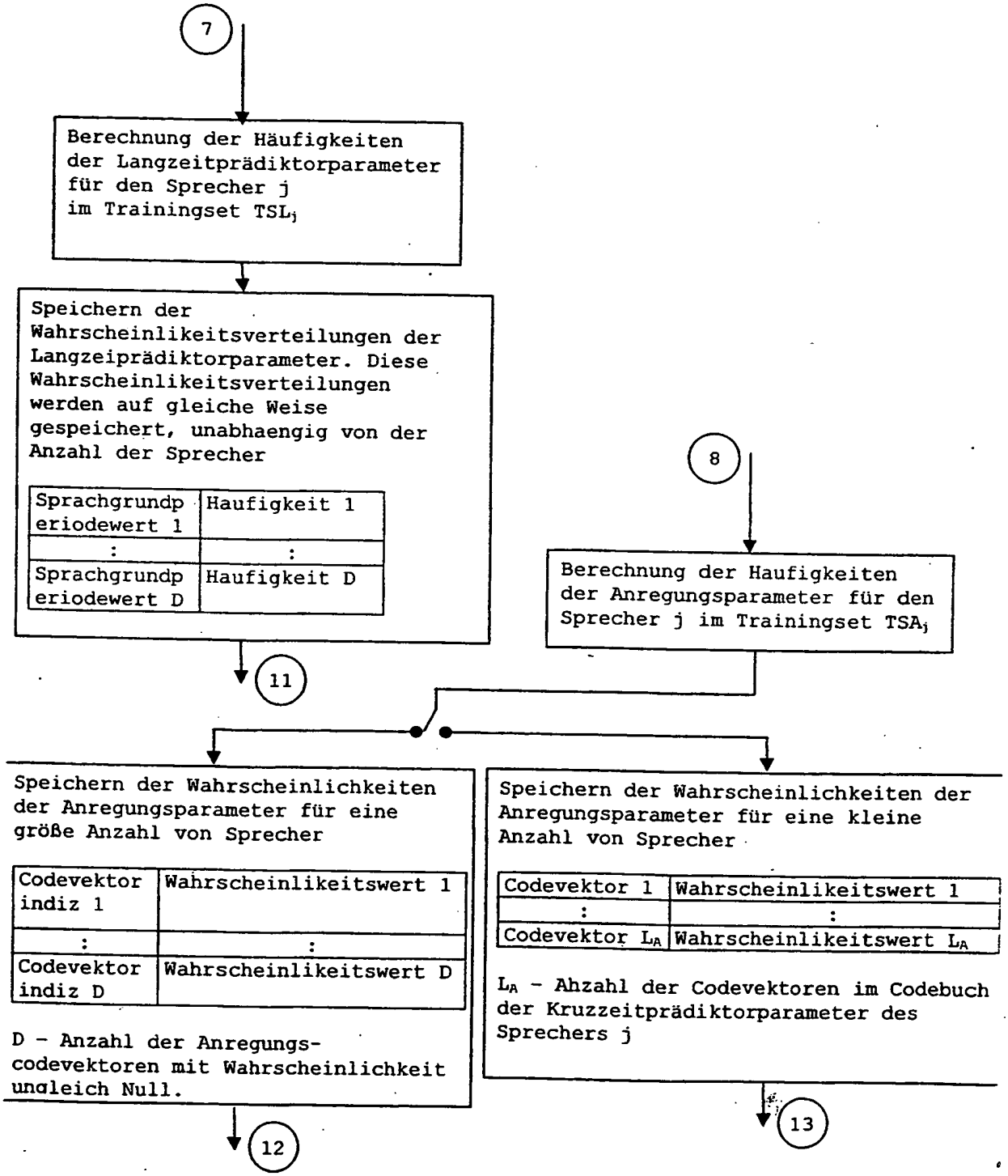
* Der im folgenden defienierter Prozess wird, für jeden neuen Nutzer des Sprechererkennungssystems durchgeführt. Der Ziel der Vorbereitungsphase ist die Erstellung der Sprecherdaten für jeden der M Sprecher.

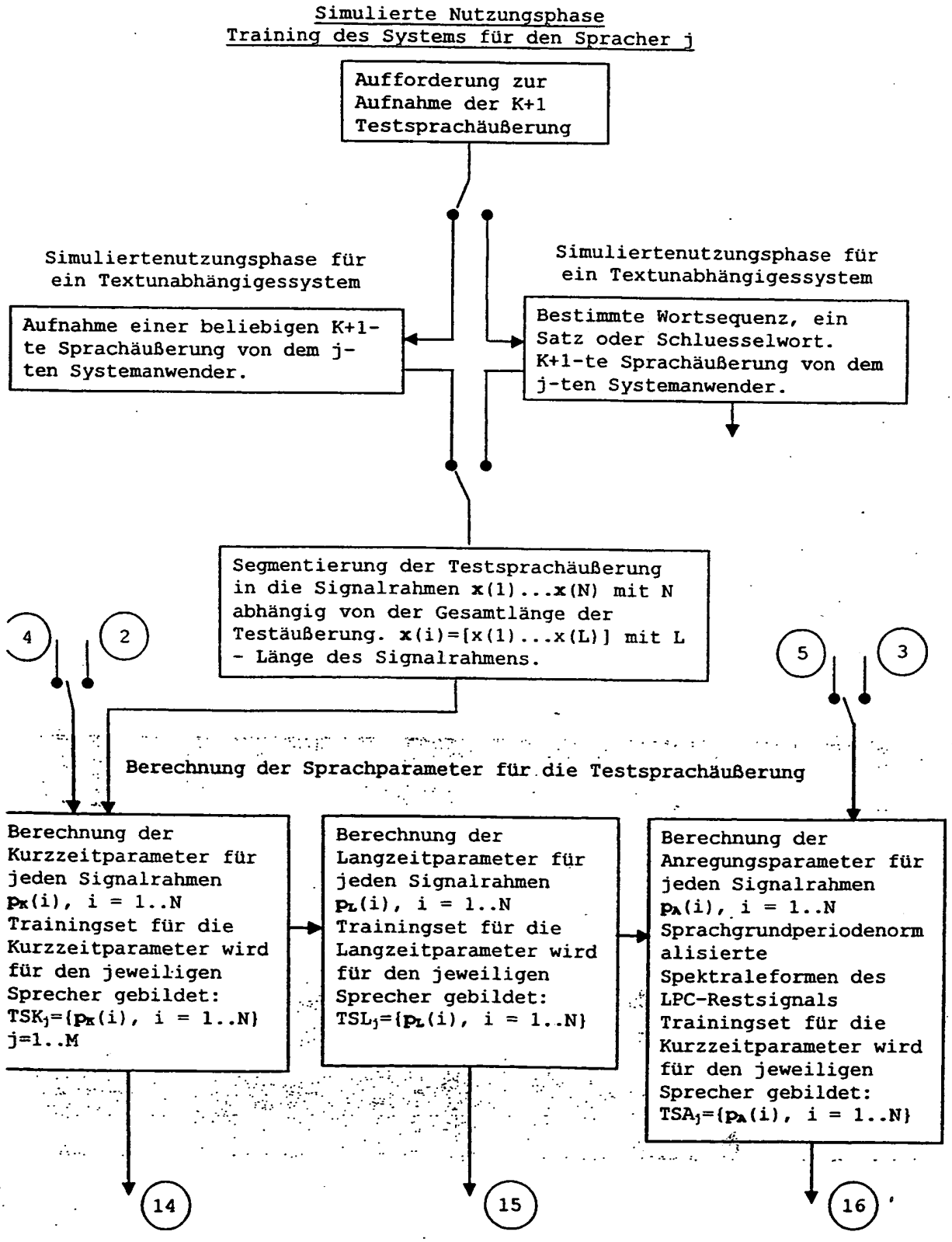


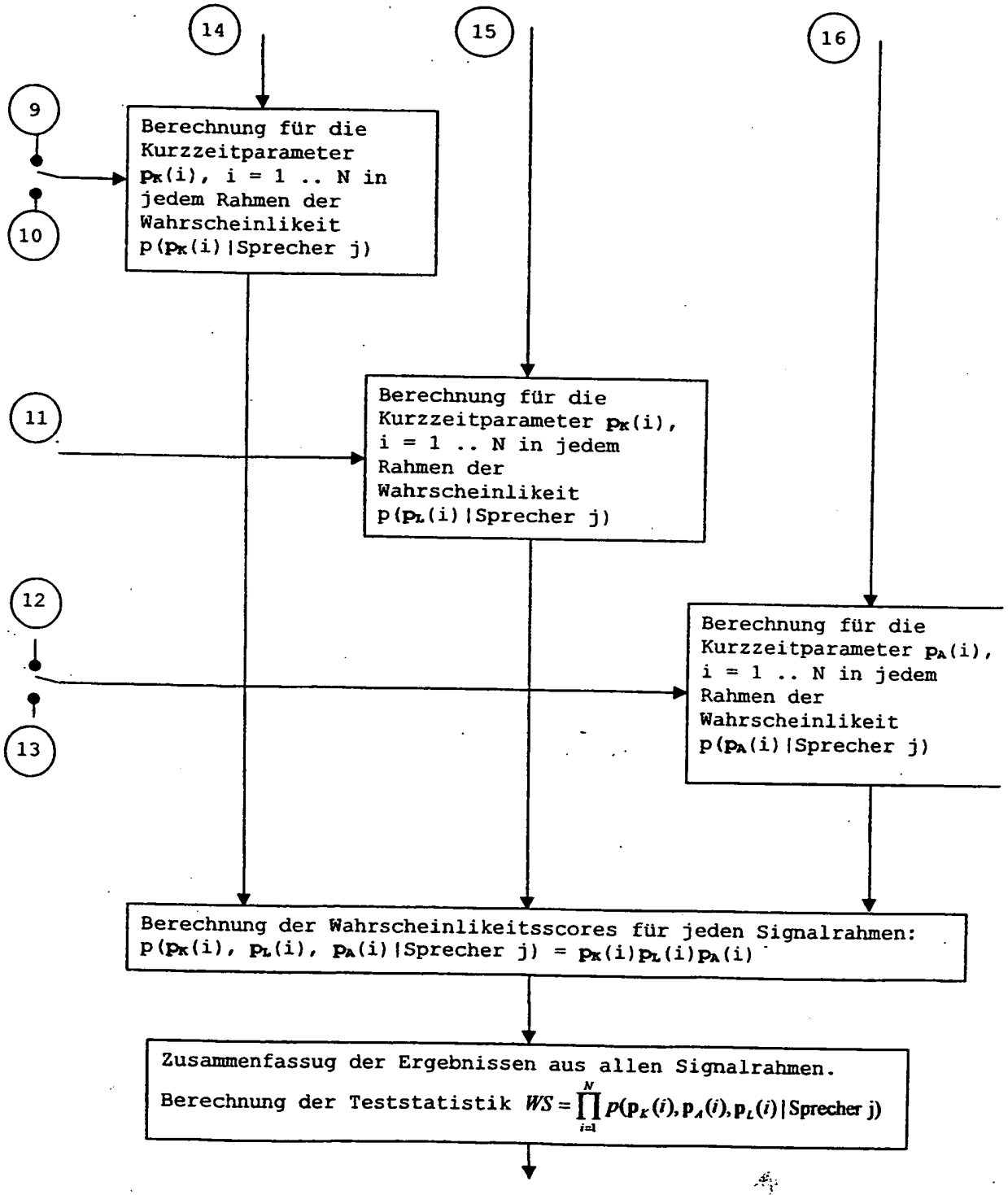
* ISO/IEC 14496-3 Information Technology - Very Low Bitrate Audio-Visual Coding

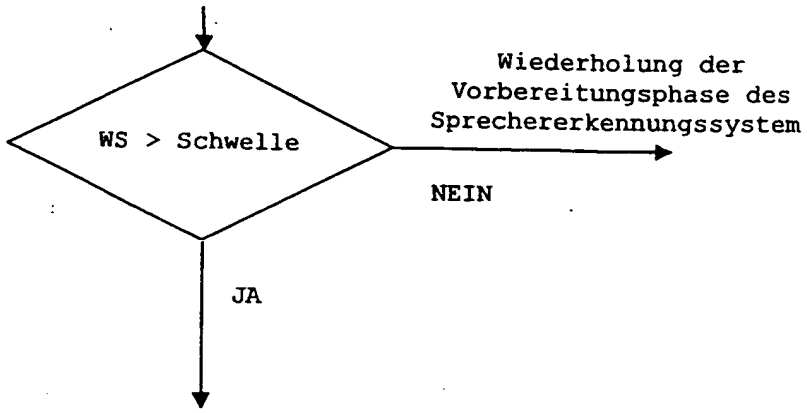
Berechnung der Volumen der Voronoizellenregionen für die Warscheinlichkeitsdichteschätzug für die Kurzzeitprädiktorparameter



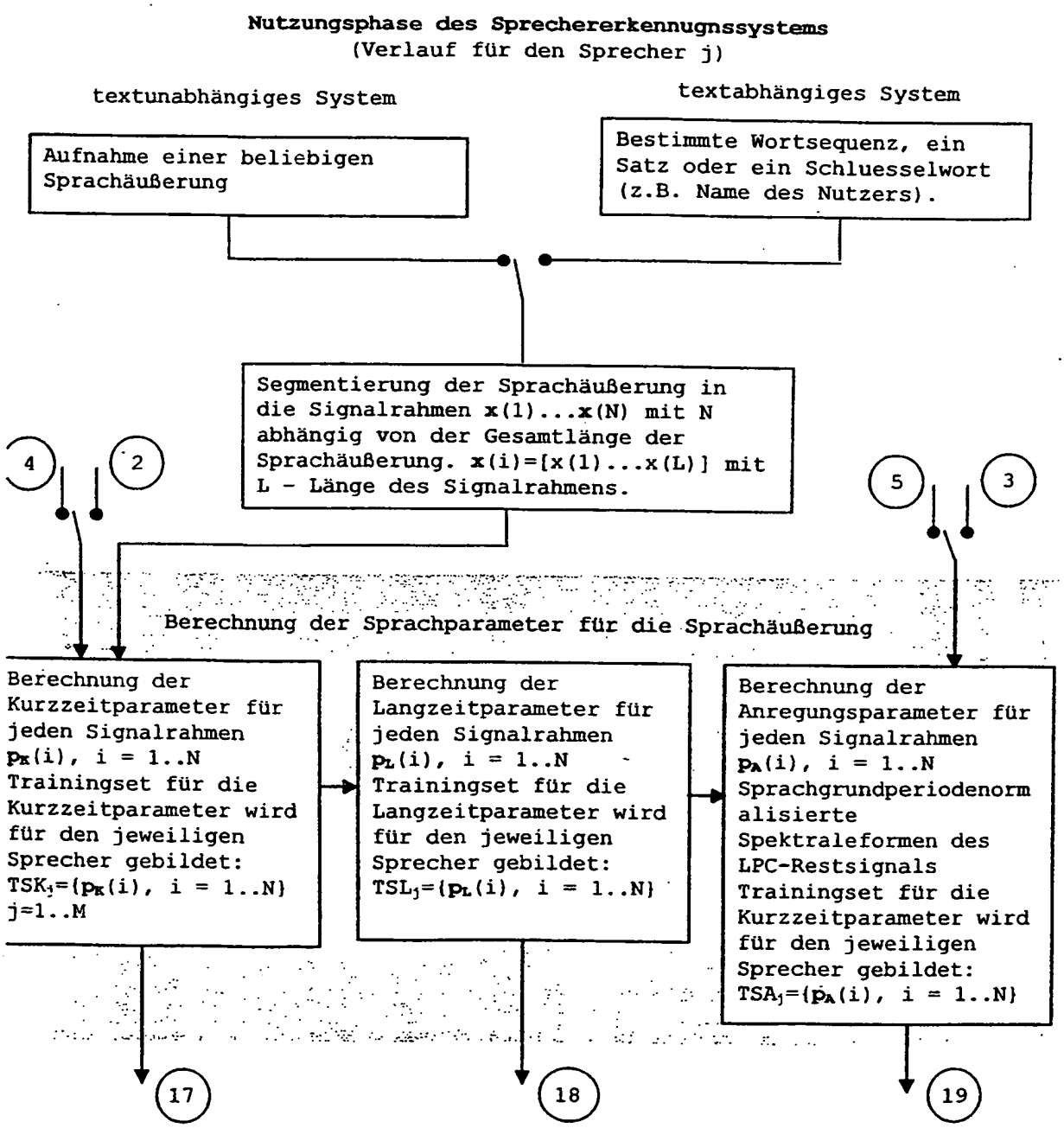

7
Berechnung der Häufigkeiten der Langzeitprädiktorparameter für den Sprecher j im Trainingset TSL$_j$
Speichern der Wahrscheinlikeitsverteilungen der Langzeiprädiktorparameter. Diese Wahrscheinlikeitsverteilungen werden auf gleiche Weise gespeichert, unabhaengig von der Anzahl der Sprecher
Sprachgrundperiodewert 1 | Haufigkeit 1
: | :
Sprachgrundperiodewert D | Haufigkeit D
11
8
Berechnung der Haufigkeiten der Anregungsparameter für den Sprecher j im Trainingset TSA$_j$
Speichern der Wahrscheinlichkeiten der Anregungsparameter für eine größe Anzahl von Sprecher
Codevektor indiz 1 | Wahrscheinlikeitswert 1
: | :
Codevektor indiz D | Wahrscheinlikeitswert D
D - Anzahl der Anregungs-codevektoren mit Wahrscheinlichkeit ungleich Null.
12
Speichern der Wahrscheinlichkeiten der Anregungsparameter für eine kleine Anzahl von Sprecher
Codevektor 1 | Wahrscheinlikeitswert 1
: | :
Codevektor L_A | Wahrscheinlikeitswert L_A
L_A - Ahzahl der Codevektoren im Codebuch der Kruzzeitprädiktorparameter des Sprechers j
13

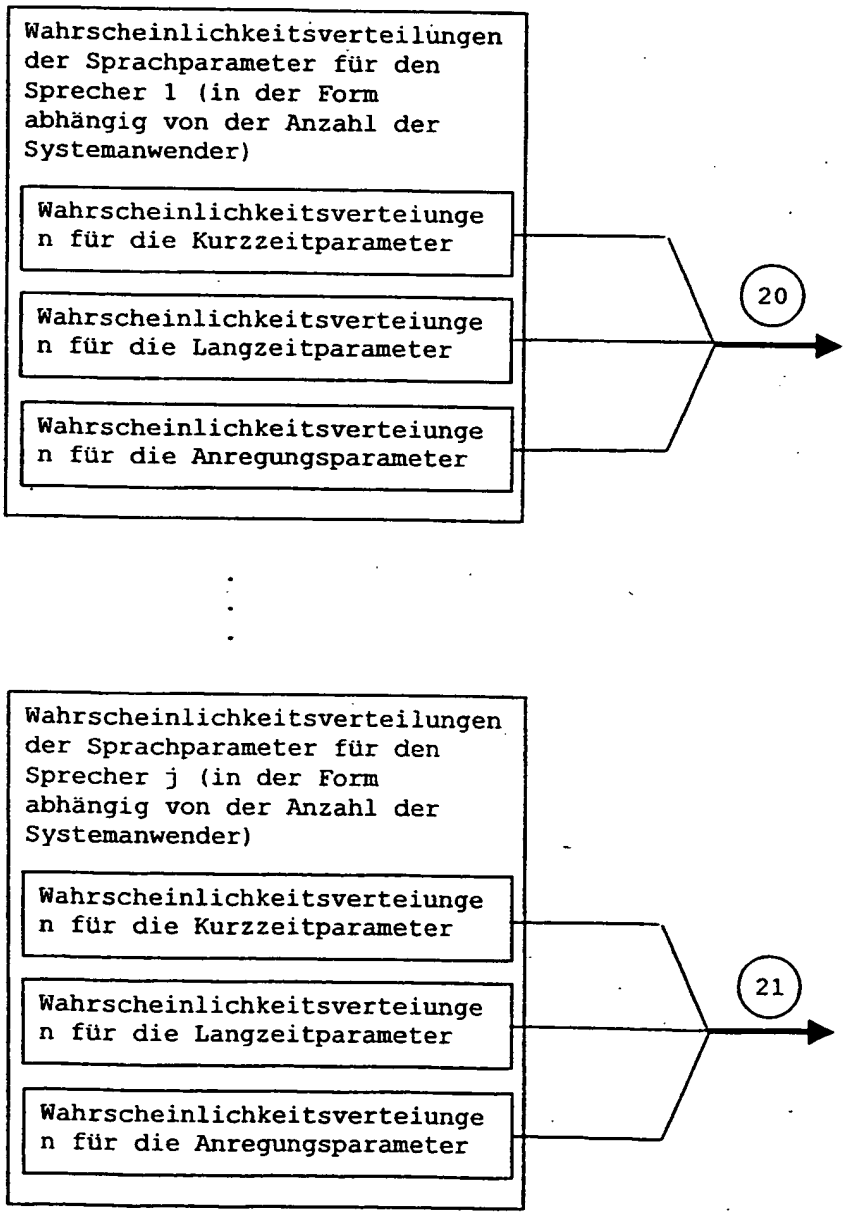

Simulierte Nutzungsphase
Training des Systems für den Spracher j
Aufforderung zur Aufnahme der K+1 Testsprachäußerung
Simuliertenutzungsphase für ein Textunabhängigessystem
Aufnahme einer beliebigen K+1-te Sprachäußerung von dem j-ten Systemanwender.
Simuliertenutzungsphase für ein Textunabhängigessystem
Bestimmte Wortsequenz, ein Satz oder Schluesselwort. K+1-te Sprachäußerung von dem j-ten Systemanwender.
Segmentierung der Testsprachäußerung in die Signalrahmen x(1)...x(N) mit N abhängig von der Gesamtlänge der Testäußerung. x(i)=[x(1)...x(L)] mit L - Länge des Signalrahmens.
4
2
5
3
Berechnung der Sprachparameter für die Testsprachäußerung
Berechnung der Kurzzeitparameter für jeden Signalrahmen $p_K(i)$, i = 1..N Trainingset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: $TSK_j=\{p_K(i), i = 1..N\}$ j=1..M
Berechnung der Langzeitparameter für jeden Signalrahmen $p_L(i)$, i = 1..N Trainingset für die Langzeitparameter wird für den jeweiligen Sprecher gebildet: $TSL_j=\{p_L(i), i = 1..N\}$
Berechnung der Anregungsparameter für jeden Signalrahmen $p_A(i)$, i = 1..N Sprachgrundperiodenormalisierte Spektraleformen des LPC-Restsignals Trainingset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: $TSA_j=\{p_A(i), i = 1..N\}$
14
15
16

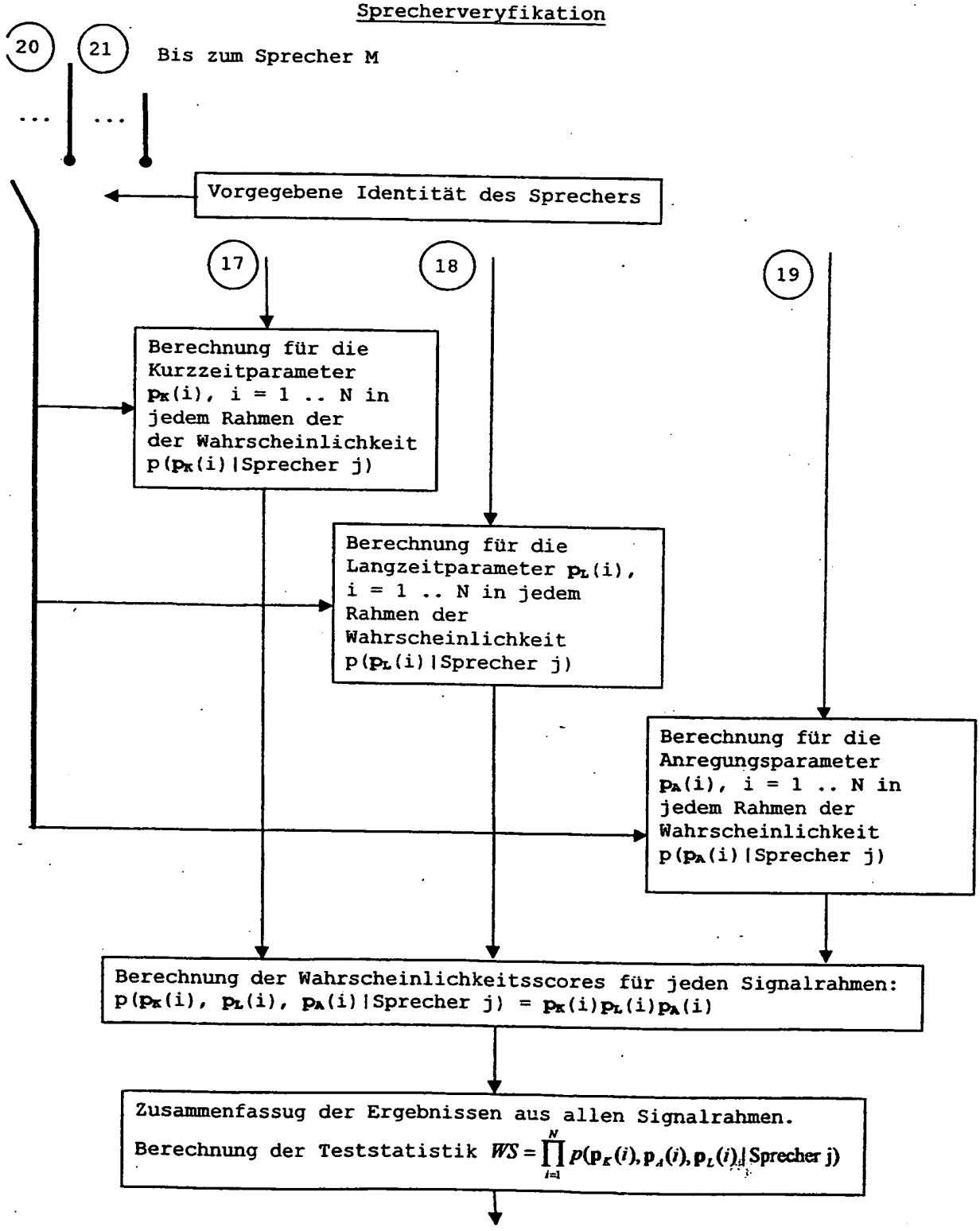

14
15
16
9
10
11
12
13
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlikeit $p(p_K(i)|\text{Sprecher j})$
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlikeit $p(p_L(i)|\text{Sprecher j})$
Berechnung für die Kurzzeitparameter $p_A(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlikeit $p(p_A(i)|\text{Sprecher j})$
Berechnung der Wahrscheinlikeitsscores für jeden Signalrahmen: $p(p_K(i), p_L(i), p_A(i)|\text{Sprecher j}) = p_K(i)p_L(i)p_A(i)$
Zusammenfassug der Ergebnissen aus allen Signalrahmen.
Berechnung der Teststatistik $WS = \prod_{i=1}^{N} p(p_K(i), p_A(i), p_L(i)|\text{Sprecher j})$

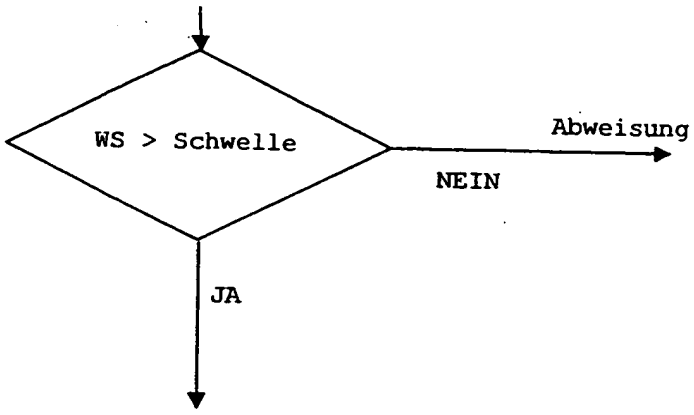


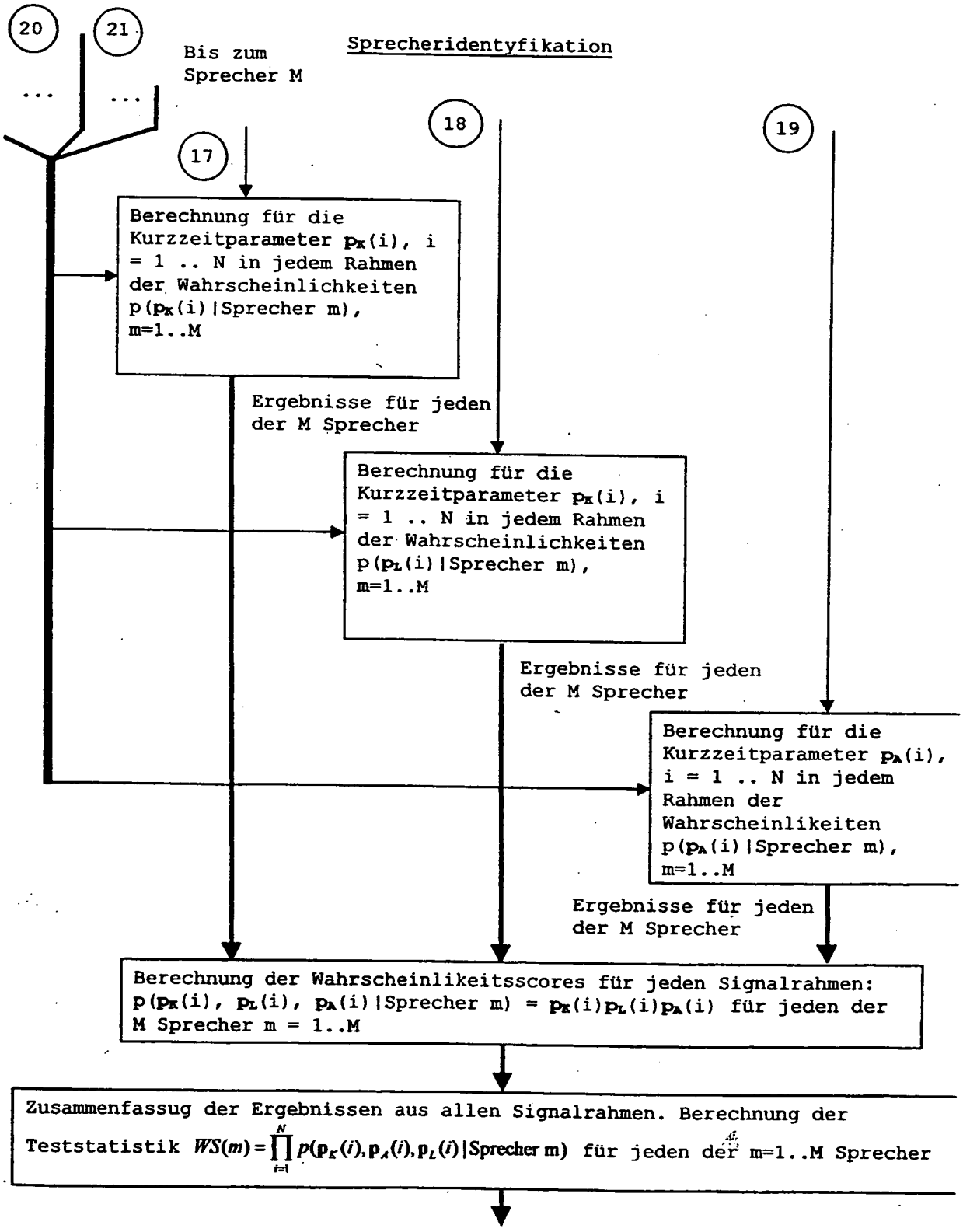
kein zusätzliches Training der Wahrscheinlichkeitsverteilungen nötig. Die Wahrscheinlichkeitsverteilungen werden im System gespeichert und sind fertig für die Nutzungphase.

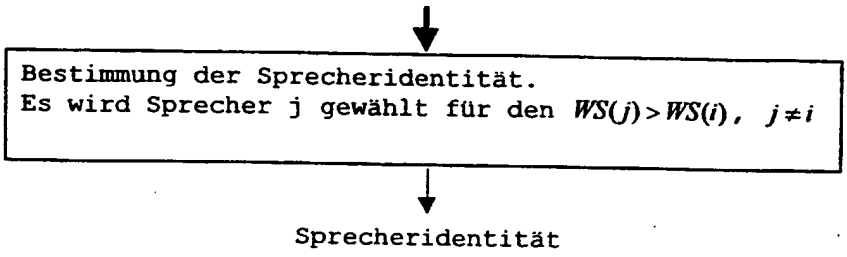

Nutzungsphase des Sprechererkennugnssystems
(Verlauf für den Sprecher j)
textunabhängiges System
textabhängiges System
Aufnahme einer beliebigen Sprachäußerung
Bestimmte Wortsequenz, ein Satz oder ein Schluesselwort (z.B. Name des Nutzers).
Segmentierung der Sprachäußerung in die Signalrahmen x(1)...x(N) mit N abhängig von der Gesamtlänge der Sprachäußerung. x(i)=[x(1)...x(L)] mit L - Länge des Signalrahmens.
4
2
5
3
Berechnung der Sprachparameter für die Sprachäußerung
Berechnung der Kurzzeitparameter für jeden Signalrahmen pK(i), i = 1..N Trainingset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: TSKj={pK(i), i = 1..N} j=1..M
Berechnung der Langzeitparameter für jeden Signalrahmen pL(i), i = 1..N Trainingset für die Langzeitparameter wird für den jeweiligen Sprecher gebildet: TSLj={pL(i), i = 1..N}
Berechnung der Anregungsparameter für jeden Signalrahmen pA(i), i = 1..N Sprachgrundperiodenormalisierte Spektraleformen des LPC-Restsignals Trainingset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: TSAj={pA(i), i = 1..N}
17
18
19


Wahrscheinlichkeitsverteilungen der Sprachparameter für den Sprecher 1 (in der Form abhängig von der Anzahl der Systemanwender)
Wahrscheinlichkeitsverteiunge n für die Kurzzeitparameter
Wahrscheinlichkeitsverteiunge n für die Langzeitparameter
Wahrscheinlichkeitsverteiunge n für die Anregungsparameter
20
Wahrscheinlichkeitsverteilungen der Sprachparameter für den Sprecher j (in der Form abhängig von der Anzahl der Systemanwender)
Wahrscheinlichkeitsverteiunge n für die Kurzzeitparameter
Wahrscheinlichkeitsverteiunge n für die Langzeitparameter
Wahrscheinlichkeitsverteiunge n für die Anregungsparameter
21
bis zum Sprecher M


Sprecherveryfikation
20
21
Bis zum Sprecher M
...
...
Vorgegebene Identität des Sprechers
17
18
19
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der der Wahrscheinlichkeit $p(p_K(i)|\text{Sprecher j})$
Berechnung für die Langzeitparameter $p_L(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_L(i)|\text{Sprecher j})$
Berechnung für die Anregungsparameter $p_A(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_A(i)|\text{Sprecher j})$
Berechnung der Wahrscheinlichkeitsscores für jeden Signalrahmen: $p(p_K(i), p_L(i), p_A(i)|\text{Sprecher j}) = p_K(i)p_L(i)p_A(i)$
Zusammenfassug der Ergebnissen aus allen Signalrahmen.
Berechnung der Teststatistik $WS = \prod_{i=1}^{N} p(p_K(i), p_A(i), p_L(i)|\text{Sprecher j})$


WS > Schwelle
Abweisung
NEIN
JA
Bestätigung der Sprecheridentität


20
21
Bis zum Sprecher M
Sprecheridentyfikation
18
19
17
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeiten $p(p_K(i)|\text{Sprecher m})$, m=1..M
Ergebnisse für jeden der M Sprecher
Berechnung für die Kurzzeitparameter $p_L(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeiten $p(p_L(i)|\text{Sprecher m})$, m=1..M
Ergebnisse für jeden der M Sprecher
Berechnung für die Kurzzeitparameter $p_A(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlikeiten $p(p_A(i)|\text{Sprecher m})$, m=1..M
Ergebnisse für jeden der M Sprecher
Berechnung der Wahrscheinlikeitsscores für jeden Signalrahmen: $p(p_K(i), p_L(i), p_A(i)|\text{Sprecher m}) = p_K(i)p_L(i)p_A(i)$ für jeden der M Sprecher m = 1..M
Zusammenfassug der Ergebnissen aus allen Signalrahmen. Berechnung der Teststatistik $WS(m) = \prod_{i=1}^{N} p(p_K(i), p_A(i), p_L(i)|\text{Sprecher m})$ für jeden der m=1..M Sprecher


Bestimmung der Sprecheridentität.
Es wird Sprecher j gewählt für den $WS(j) > WS(i),\ j \neq i$
Sprecheridentität

INTERNATIONAL SEARCH REPORT

Inter application No
PCT/DE 00/02917

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G10L17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	MOGAKI T ET AL: "Text-indicated speaker verification method using PSI-CELP parameters" SECURITY AND WATERMARKING OF MULTIMEDIA CONTENTS, SAN JOSE, CA, USA, 25-27 JAN. 1999, vol. 3657, pages 184-193, XP000981232 Proceedings of the SPIE - The International Society for Optical Engineering, 1999, SPIE-Int. Soc. Opt. Eng, USA ISSN: 0277-786X page 2 figure 5	1,3-6

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 January 2001	09/02/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Krembel, L

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 0 817 170 A (TELIA AB) 7 January 1998 (1998-01-07) column 2, line 15 - line 22 column 4, line 1 - line 15	1,3
Y	US 5 535 305 A (CHOW YEN-LU ET AL) 9 July 1996 (1996-07-09) column 1, line 6 - line 12	4
Y	BOLL S F: "SUPPRESSION OF ACOUSTIC NOISE IN SPEECH USING SPECTRAL SUBTRACTION" IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING,US,IEEE INC. NEW YORK, vol. 27, no. 2, 1 April 1979 (1979-04-01), pages 113-120, XP000560467 abstract	5,6

INTERNATIO L SEARCH REPORT

...formation on patent family members

Inter Application No
PCT/DE 00/02917

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0817170 A	07-01-1998	SE 505522 C	08-09-1997
		US 5960392 A	28-09-1999
		NO 972670 A	02-01-1998
		SE 9602622 A	08-09-1997
US 5535305 A	09-07-1996	NONE	

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS

Absender: INTERNATIONALE RECHERCHENBEHÖRDE

PCT

MITTEILUNG ÜBER DIE ÜBERMITTLUNG DES INTERNATIONALEN RECHERCHENBERICHTS ODER DER ERKLÄRUNG

(Regel 44.1 PCT)

An
SIEMENS AKTIENGESELLSCHAFT
Postfach 22 16 34
80506 München
GERMANY

ZT GG ViM Mch P/Ri
Eing. 12. Feb. 2001
GR Frist 26.03.01

Aktenzeichen des Anmelders oder Anwalts 1999P02665WO	Absendedatum *(Tag/Monat/Jahr)* 09/02/2001 **WEITERES VORGEHEN** siehe Punkte 1 und 4 unten
Internationales Aktenzeichen PCT/DE 00/ 02917	Internationales Anmeldedatum *(Tag/Monat/Jahr)* 25/08/2000
Anmelder SIEMENS AKTIENGESELLSCHAFT et al.	

1. [X] Dem Anmelder wird mitgeteilt, daß der internationale Recherchenbericht erstellt wurde und ihm hiermit übermittelt wird.

Einreichung von Änderungen und einer Erklärung nach Artikel 19:
Der Anmelder kann auf eigenen Wunsch die Ansprüche der internationalen Anmeldung ändern (siehe Regel 46):

Bis wann sind Änderungen einzureichen?

Die Frist zur Einreichung solcher Änderungen beträgt üblicherweise zwei Monate ab der Übermittlung des internationalen Recherchenberichts; weitere Einzelheiten sind den Anmerkungen auf dem Beiblatt zu entnehmen.

Wo sind Änderungen einzureichen?

Unmittelbar beim Internationalen Büro der WIPO, 34, CHEMIN des Colombettes, CH-1211 Genf 20, Telefaxnr.: (41-22) 740.14.35

Nähere Hinweise sind den Anmerkungen auf dem Beiblatt zu entnehmen.

2. [ ] Dem Anmelder wird mitgeteilt, daß kein internationaler Recherchenbericht erstellt wird und daß ihm hiermit die Erklärung nach Artikel 17(2)a) übermittelt wird.

3. [ ] **Hinsichtlich des Widerspruchs** gegen die Entrichtung einer zusätzlichen Gebühr (zusätzlicher Gebühren) nach Regel 40.2 wird dem Anmelder mitgeteilt, daß

[ ] der Widerspruch und die Entscheidung hierüber zusammen mit seinem Antrag auf Übermittlung des Wortlauts sowohl des Widerspruchs als auch der Entscheidung hierüber an die Bestimmungsämter dem Internationalen Büro übermittelt worden sind.

[ ] noch keine Entscheidung über den Widerspruch vorliegt; der Anmelder wird benachrichtigt, sobald eine Entscheidung getroffen wurde.

4. **Weiteres Vorgehen:** Der Anmelder wird auf folgendes aufmerksam gemacht:

Kurz nach Ablauf von **18 Monaten** seit dem Prioritätsdatum wird die internationale Anmeldung vom Internationalen Büro veröffentlicht. Will der Anmelder die Veröffentlichung verhindern oder auf einen späteren Zeitpunkt verschieben, so muß gemäß Regel 90 bis.1 bzw. 90 bis.3 vor Abschluß der technischen Vorbereitungen für die internationale Veröffentlichung eine Erklärung über die Zurücknahme der internationalen Anmeldung oder des Prioritätsanspruchs beim Internationalen Büro eingehen.

Innerhalb von **19 Monaten** seit dem Prioritätsdatum ist ein Antrag auf internationale vorläufige Prüfung einzureichen, wenn der Anmelder den Eintritt in die nationale Phase bis zu 30 Monaten seit dem Prioritätsdatum (in manchen Ämtern sogar noch länger) verschieben möchte.

Innerhalb von **20 Monaten** seit dem Prioritätsdatum muß der Anmelder die für den Eintritt in die nationale Phase vorgeschriebenen Handlungen vor allen Bestimmungsämtern vornehmen, die nicht innerhalb von 19 Monaten seit dem Prioritätsdatum in der Anmeldung oder einer nachträglichen Auswahlerklärung ausgewählt wurden oder nicht ausgewählt werden konnten, da für sie Kapitel II des Vertrages nicht verbindlich ist.

Name und Postanschrift der Internationalen Recherchenbehörde Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL-2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Bevollmächtigter Bediensteter Ahmed Soliman

(Siehe Anmerkungen auf Beiblatt)

ANMERKUNGEN ZU FORMBLATT PCT/ISA/220

Diese Anmerkungen sollen grundlegende Hinweise zur Einreichung von Änderungen gemäß Artikel 19 geben. Diesen Anmerkungen liegen die Erfordernisse des Vertrags über die internationale Zusammenarbeit auf dem Gebiet des Patentwesens (PCT), der Ausführungsordnung und der Verwaltungsrichtlinien zu diesem Vertrag zugrunde. Bei Abweichungen zwischen diesen Anmerkungen und obengenannten Texten sind letztere maßgebend. Nähere Einzelheiten sind dem PCT-Leitfaden für Anmelder, einer Veröffentlichung der WIPO, zu entnehmen.
Die in diesen Anmerkungen verwendeten Begriffe "Artikel", "Regel" und "Abschnitt" beziehen sich jeweils auf die Bestimmungen des PCT-Vertrags, der PCT-Ausführungsordnung bzw. der PCT-Verwaltungsrichtlinien.

HINWEISE ZU ÄNDERUNGEN GEMÄSS ARTIKEL 19

Nach Erhalt des internationalen Recherchenberichts hat der Anmelder die Möglichkeit, einmal die Ansprüche der internationalen Anmeldung zu ändern. Es ist jedoch zu betonen, daß, da alle Teile der internationalen Anmeldung (Ansprüche, Beschreibung und Zeichnungen) während des internationalen vorläufigen Prüfungsverfahrens geändert werden können, normalerweise keine Notwendigkeit besteht, Änderungen der Ansprüche nach Artikel 19 einzureichen, außer wenn der Anmelder z.B. zum Zwecke eines vorläufigen Schutzes die Veröffentlichung dieser Ansprüche wünscht oder ein anderer Grund für eine Änderung der Ansprüche vor ihrer internationalen Veröffentlichung vorliegt. Weiterhin ist zu beachten, daß ein vorläufiger Schutz nur in einigen Staaten erhältlich ist.

Welche Teile der internationalen Anmeldung können geändert werden?

Im Rahmen von Artikel 19 können nur die Ansprüche geändert werden.

In der internationalen Phase können die Ansprüche auch nach Artikel 34 vor der mit der internationalen vorläufigen Prüfung beauftragten Behörde geändert (oder nochmals geändert) werden. Die Beschreibung und die Zeichnungen können nur nach Artikel 34 vor der mit der internationalen vorläufigen Prüfung beauftragten Behörde geändert werden.

Beim Eintritt in die nationale Phase können alle Teile der internationalen Anmeldung nach Artikel 28 oder gegebenenfalls Artikel 41 geändert werden.

Bis wann sind Änderungen einzureichen?

Innerhalb von zwei Monaten ab der Übermittlung des internationalen Recherchenberichts oder innerhalb von sechzehn Monaten ab dem Prioritätsdatum, je nachdem, welche Frist später abläuft. Die Änderungen gelten jedoch als rechtzeitig eingereicht, wenn sie dem Internationalen Büro nach Ablauf der maßgebenden Frist, aber noch vor Abschluß der technischen Vorbereitungen für die internationale Veröffentlichung (Regel 46.1) zugehen.

Wo sind die Änderungen nicht einzureichen?

Die Änderungen können nur beim Internationalen Büro, nicht aber beim Anmeldeamt oder der Internationalen Recherchenbehörde eingereicht werden (Regel 46.2).

Falls ein Antrag auf internationale vorläufige Prüfung eingereicht wurde/wird, siehe unten.

In welcher Form können Änderungen erfolgen?

Eine Änderung kann erfolgen durch Streichung eines oder mehrerer ganzer Ansprüche, durch Hinzufügung eines oder mehrerer neuer Ansprüche oder durch Änderung des Wortlauts eines oder mehrerer Ansprüche in der eingereichten Fassung.

Für jedes Anspruchsblatt, das sich aufgrund einer oder mehrerer Änderungen von dem ursprünglich eingereichten Blatt unterscheidet, ist ein Ersatzblatt einzureichen.

Alle Ansprüche, die auf einem Ersatzblatt erscheinen, sind mit arabischen Ziffern zu numerieren. Wird ein Anspruch gestrichen, so brauchen, die anderen Ansprüche nicht neu numeriert zu werden. Im Fall einer Neunumerierung sind die Ansprüche fortlaufend zu numerieren (Verwaltungsrichtlinien, Abschnitt 205 b)).

Die Änderungen sind in der Sprache abzufassen, in der dieinternationale Anmeldung veröffentlicht wird.

Welche Unterlagen sind den Änderungen beizufügen?

Begleitschreiben (Abschnitt 205 b)):

Die Änderungen sind mit einem Begleitschreiben einzureichen.

Das Begleitschreiben wird nicht zusammen mit der internationalen Anmeldung und den geänderten Ansprüchen veröffentlicht. Es ist nicht zu verwechseln mit der "Erklärung nach Artikel 19(1)" (siehe unten, "Erklärung nach Artikel 19 (1)").

Das Begleitschreiben ist nach Wahl des Anmelders in englischer oder französischer Sprache abzufassen. Bei englischsprachigen internationalen Anmeldungen ist das Begleitschreiben aber ebenfalls in englischer, bei französischsprachigen internationalen Anmeldungen in französischer Sprache abzufassen.

ANMERKUNGEN ZU FORMBLATT PCT/ISA/220 (F rtsetzung)

Im Begleitschreiben sind die Unterschiede zwischen den Ansprüchen in der eingereichten Fassung und den geänderten Ansprüchen anzugeben. So ist insbesondere zu jedem Anspruch in der internationalen Anmeldung anzugeben (gleichlautende Angaben zu verschiedenen Ansprüchen können zusammengefaßt werden), ob

i) der Anspruch unverändert ist;

ii) der Anspruch gestrichen worden ist;

iii) der Anspruch neu ist;

iv) der Anspruch einen oder mehrere Ansprüche in der eingereichten Fassung ersetzt;

v) der Anspruch auf die Teilung eines Anspruchs in der eingereichten Fassung zurückzuführen ist.

Im folgenden sind Beispiele angegeben, wie Änderungen im Begleitschreiben zu erläutern sind:

1. [Wenn anstelle von ursprünglich 48 Ansprüchen nach der Änderung einiger Ansprüche 51 Ansprüche existieren]:
"Die Ansprüche 1 bis 29, 31, 32, 34, 35, 37 bis 48 werden durch geänderte Ansprüche gleicher Numerierung ersetzt; Ansprüche 30, 33 und 36 unverändert; neue Ansprüche 49 bis 51 hinzugefügt."

2. [Wenn anstelle von ursprünglich 15 Ansprüchen nach der Änderung aller Ansprüche 11 Ansprüche existieren]:
"Geänderte Ansprüche 1 bis 11 treten an die Stelle der Ansprüche 1 bis 15."

3. [Wenn ursprünglich 14 Ansprüche existierten und die Änderungen darin bestehen, daß einige Ansprüche gestrichen werden und neue Ansprüche hinzugefügt werden]:
Ansprüche 1 bis 6 und 14 unverändert; Ansprüche 7 bis 13 gestrichen; neue Ansprüche 15, 16 und 17 hinzugefügt."Oder" Ansprüche 7 bis 13 gestrichen; neue Ansprüche 15, 16 und 17 hinzugefügt; alle übrigen Ansprüche unverändert."

4. [Wenn verschiedene Arten von Änderungen durchgeführt werden]:
"Ansprüche 1-10 unverändert; Ansprüche 11 bis 13, 18 und 19 gestrichen; Ansprüche 14, 15 und 16 durch geänderten Anspruch 14 ersetzt; Anspruch 17 in geänderte Ansprüche 15, 16 und 17 unterteilt; neue Ansprüche 20 und 21 hinzugefügt."

"Erklärung nach Artikel 19(1)" (Regel 46.4)

Den Änderungen kann eine Erklärung beigefügt werden, mit der die Änderungen erläutert und ihre Auswirkungen auf die Beschreibung und die Zeichnungen dargelegt werden (die nicht nach Artikel 19 (1) geändert werden können).

Die Erklärung wird zusammen mit der internationalen Anmeldung und den geänderten Ansprüchen veröffentlicht.

Sie ist in der Sprache abzufassen, in der die internationale Anmeldung veröffentlicht wird.

Sie muß kurz gehalten sein und darf, wenn in englischer Sprache abgefaßt oder ins Englische übersetzt, nicht mehr als 500 Wörter umfassen

Die Erklärung ist nicht zu verwechseln mit dem Begleitschreiben, das auf die Unterschiede zwischen den Ansprüchen in der eingereichten Fassung und den geänderten Ansprüchen hinweist, und ersetzt letzteres nicht. Sie ist auf einem gesonderten Blatt einzureichen und in der Überschrift als solche zu kennzeichnen, vorzugsweise mit den Worten "Erklärung nach Artikel 19 (1)".

Die Erklärung darf keine herabsetzenden Äußerungen über den internationalen Recherchenbericht oder die Bedeutung von in dem Bericht angeführten Veröffentlichungen enthalten. Sie darf auf im internationalen Recherchenbericht angeführte Veröffentlichungen, die sich auf einen bestimmten Anspruch beziehen, nur im Zusammenhang mit einer Änderung dieses Anspruchs Bezug nehmen.

Auswirkungen eines bereits gestellten Antrags auf internationale vorläufige Prüfung

Ist zum Zeitpunkt der Einreichung von Änderungen nach Artikel 19 bereits ein Antrag auf internationale vorläufige Prüfung gestellt worden, so sollte der Anmelder in seinem Interesse gleichzeitig mit der Einreichung der Änderungen beim Internationalen Büro auch eine Kopie der Änderungen bei der mit der internationalen vorläufigen Prüfung beauftragen Behörde einreichen (siehe Regel 62.2 a), erster Satz).

Auswirkungen von Änderungen hinsichtlich der Übersetzung der internationalen Anmeldung beim Eintritt in die nationale Phase

Der Anmelder wird darauf hingewiesen, daß bei Eintritt in die nationale Phase möglicherweise anstatt oder zusätzlich zu der Übersetzung der Ansprüche in der eingereichten Fassung eine Übersetzung der nach Artikel 19 geänderten Ansprüche an die bestimmten/ausgewählten Ämter zu übermitteln ist.

Nähere Einzelheiten über die Erfordernisse jedes bestimmten/ausgewählten Amts sind Band II des PCT-Leitfadens für Anmelder zu entnehmen.

VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS

PCT

INTERNATIONALER RECHERCHENBERICHT

(Artikel 18 sowie Regeln 43 und 44 PCT)

Aktenzeichen des Anmelders oder Anwalts	**WEITERES VORGEHEN**	siehe Mitteilung über die Übermittlung des internationalen Recherchenberichts (Formblatt PCT/ISA/220) sowie, soweit zutreffend, nachstehender Punkt 5
1999P02665WO		
Internationales Aktenzeichen	Internationales Anmeldedatum *(Tag/Monat/Jahr)*	(Frühestes) Prioritätsdatum *(Tag/Monat/Jahr)*
PCT/DE 00/ 02917	25/08/2000	26/08/1999
Anmelder		
SIEMENS AKTIENGESELLSCHAFT et al.		

Dieser internationale Recherchenbericht wurde von der Internationalen Recherchenbehörde erstellt und wird dem Anmelder gemäß Artikel 18 übermittelt. Eine Kopie wird dem Internationalen Büro übermittelt.

Dieser internationale Recherchenbericht umfaßt insgesamt 3 Blätter.

[X] Darüber hinaus liegt ihm jeweils eine Kopie der in diesem Bericht genannten Unterlagen zum Stand der Technik bei.

1. **Grundlage des Berichts**

   a. Hinsichtlich der **Sprache** ist die internationale Recherche auf der Grundlage der internationalen Anmeldung in der Sprache durchgeführt worden, in der sie eingereicht wurde, sofern unter diesem Punkt nichts anderes angegeben ist.

   [ ] Die internationale Recherche ist auf der Grundlage einer bei der Behörde eingereichten Übersetzung der internationalen Anmeldung (Regel 23.1 b)) durchgeführt worden.

   b. Hinsichtlich der in der internationalen Anmeldung offenbarten **Nucleotid– und/oder Aminosäuresequenz** ist die internationale Recherche auf der Grundlage des Sequenzprotokolls durchgeführt worden, das

   [ ] in der internationalen Anmeldung in Schriftlicher Form enthalten ist.

   [ ] zusammen mit der internationalen Anmeldung in computerlesbarer Form eingereicht worden ist.

   [ ] bei der Behörde nachträglich in schriftlicher Form eingereicht worden ist.

   [ ] bei der Behörde nachträglich in computerlesbarer Form eingereicht worden ist.

   [ ] Die Erklärung, daß das nachträglich eingereichte schriftliche Sequenzprotokoll nicht über den Offenbarungsgehalt der internationalen Anmeldung im Anmeldezeitpunkt hinausgeht, wurde vorgelegt.

   [ ] Die Erklärung, daß die in computerlesbarer Form erfaßten Informationen dem schriftlichen Sequenzprotokoll entsprechen, wurde vorgelegt.

2. [ ] **Bestimmte Ansprüche haben sich als nicht recherchierbar erwiesen** (siehe Feld I).

3. [ ] **Mangelnde Einheitlichkeit der Erfindung** (siehe Feld II).

4. Hinsichtlich der **Bezeichnung der Erfindung**

   [ ] wird der vom Anmelder eingereichte Wortlaut genehmigt.

   [X] wurde der Wortlaut von der Behörde wie folgt festgesetzt:

   VERFAHREN ZUM TRAINIEREN EINES SPRECHERERKENNUNGSSYSTEMS

5. Hinsichtlich der **Zusammenfassung**

   [X] wird der vom Anmelder eingereichte Wortlaut genehmigt.

   [ ] wurde der Wortlaut nach Regel 38.2b) in der in Feld III angegebenen Fassung von der Behörde festgesetzt. Der Anmelder kann der Behörde innerhalb eines Monats nach dem Datum der Absendung dieses internationalen Recherchenberichts eine Stellungnahme vorlegen.

6. Folgende Abbildung der **Zeichnungen** ist mit der Zusammenfassung zu veröffentlichen: Abb. Nr. 1

   [ ] wie vom Anmelder vorgeschlagen [ ] keine der Abb.

   [X] weil der Anmelder selbst keine Abbildung vorgeschlagen hat.

   [ ] weil diese Abbildung die Erfindung besser kennzeichnet.

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen
PC[redacted]E 00/02917

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES
IPK 7 G10L17/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)
IPK 7 G10L

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)
EPO-Internal, WPI Data, PAJ, INSPEC

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y ✓	MOGAKI T ET AL: "Text-indicated speaker verification method using PSI-CELP parameters" SECURITY AND WATERMARKING OF MULTIMEDIA CONTENTS, SAN JOSE, CA, USA, 25-27 JAN. 1999, Bd. 3657, Seiten 184-193, XP000981232 Proceedings of the SPIE - The International Society for Optical Engineering, 1999, SPIE-Int. Soc. Opt. Eng, USA ISSN: 0277-786X Seite 2 Abbildung 5 --- -/--	1,3-6

[X] Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

[X] Siehe Anhang Patentfamilie

° Besondere Kategorien von angegebenen Veröffentlichungen :
"A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist
"E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist
"L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)
"O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht
"P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist
"T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist
"X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden
"Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist
"&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche	Absendedatum des internationalen Recherchenberichts
26. Januar 2001	09/02/2001

Name und Postanschrift der Internationalen Recherchenbehörde	Bevollmächtigter Bediensteter
Europäisches Patentamt, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	Krembel, L

1

Formblatt PCT/ISA/210 (Blatt 2) (Juli 1992)

INTERNATIONALER RECHERCHENBERICHT

[redacted]

Internationales Aktenzeichen
PC[redacted]E 00/02917

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie°	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
Y	EP 0 817 170 A (TELIA AB) 7. Januar 1998 (1998-01-07) Spalte 2, Zeile 15 - Zeile 22 Spalte 4, Zeile 1 - Zeile 15 ---	1,3
Y	US 5 535 305 A (CHOW YEN-LU ET AL) 9. Juli 1996 (1996-07-09) Spalte 1, Zeile 6 - Zeile 12 ---	4
Y	BOLL S F: "SUPPRESSION OF ACOUSTIC NOISE IN SPEECH USING SPECTRAL SUBTRACTION" IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING,US,IEEE INC. NEW YORK, Bd. 27, Nr. 2, 1. April 1979 (1979-04-01), Seiten 113-120, XP000560467 Zusammenfassung -----	5,6

1

Angaben zu Veröffentlichungen, die zur Selben Patentfamilie gehören

Inter Aktenzeichen
PCT/DE 00/02917

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 0817170 A	07-01-1998	SE 505522 C	08-09-1997
		US 5960392 A	28-09-1999
		NO 972670 A	02-01-1998
		SE 9602622 A	08-09-1997
US 5535305 A	09-07-1996	KEINE	

09/830497
JC18 Rec'd PCT/PTO 26 APR 2001

Beschreibung

Verfahren zum Erkennen von Sprechern anhand deren Stimmen

Die Erfindung betrifft ein Verfahren zum Erkennen von Sprechern anhand deren Stimmen.

Die der Erfindung zugrundeliegende Aufgabe besteht darin, ein Verfahren zum Erkennen von Sprechern anhand deren Stimmen anzugeben, das robust, sicher und zuverlässig ist.

Diese Aufgabe wird erfindungsgemäß durch die im Patentanspruch 1 angegebenen Merkmale gelöst.

Im folgenden wird die Erfindung unter Verwendung eines Flußdiagramms näher beschrieben.

1.
Die Erfindung ermöglicht die Erkennung des Sprechers anhand seiner Stimme. Das Problem der Sprechererkennung besteht darin, zwischen verschiedenen Sprechern zu unterscheiden oder die vorgegebene Sprecheridentität zu überprüfen, wobei die einzige Eingangsinformation die Aufzeichnung der Stimme des Sprechers ist.

Außerdem wird eine Methode vorgeschlagen, die das Überlisten des Zugangssystems verhindert, wenn die Stimme und das Schlüsselwort von Dritten aufgenommen wird.

Bei der Speicherung von komplexen Wahrscheinlichkeitsverteilungen für die Sprachparameter eines Sprechers muß zwischen Genauigkeit und Speicherbedarf ein Kompromiss geschlossen werden. Deswegen werden Methoden der Speicherung der Wahrscheinlichkeitsverteilungen vorgeschlagen, die abhängig von der Anzahl der Sprecher einsetzbar sind.

2.
Die Sprechererkennung wurde bisher z.B. mit Hilfe von Hidden-Markov Modellen oder durch Vektorquantisierung gelöst, siehe Literatur [1].

3.
Die Erfindung löst das Problem der Sprechererkennung basierend auf den Parametern einer Analyse durch Synthese Kodierers mit der Linearen Prädiktion (LPAS) [1] (z.B. eines Harmonic Vector Excited Codecs [5] oder Waveform Interpolation Codec [4]). Die bisher verwendeten Parameter des Sprachsignals wie z.B. Cepstrale AR Parameter bringen keine zufriedenstellende Lösung des Problems. Deswegen muß auf andere Parameter zugegriffen werden wie z.B. Parameter der Anregung des Vokaltraktes, die sprecherabhängige und zugleich weitgehend phonemenunabhängige Information tragen.

Darüber hinaus wird die Methode der Schätzung der Wahrscheinlichkeitsverteilung der Kodiererparameter für den jeweiligen Sprecher gegeben, und eine Methode, die das Überlisten des Zugangssystems verhindert.

Sprecheridentifikation

Bei Systemen zur *Sprechererkennung* wird nach den statistischen Prinzipen [2] geprüft, ob der gesprochene Satz von einem der vom Sprechererkennungssystem erfassten Sprecher gesprochen wurde. Dabei gibt es grundsätzlich zwei Arten von Sprechererkennungssystemen, die textabhängigen und die textunabhängigen Systeme. Für die in der Erfindung beschriebene Prozedur wird die Textunabhängigkeit des System durch eine erweiterte Trainingsphase erreicht, in der der Sprecher ein vielfältiges Material aufzeichnen muß und die Wahrscheinlichkeitsverteilungen der erwähnten Sprachsignalparameter aus dem gesamten Sprachmaterial bestimmt. Das Trainieren eines textabhängigen Systems ist eine einfachere Aufgabe, weil das Sprachmaterial, das vom Sprecher während der Nutzungsphase gesprochen wird, auf einige Schlüsselworte oder bestimmte

Sätze begrenzt ist. Die Vorbereitungsphase wird so lange durchgeführt, bis das System sicher die Stimme des Sprechers erkennt.

Die Aufgabe der *Sprecheridentifikation* ist in Figur 1 (Problem der Sprecheridentifikation) dargestellt.

Die *Sprecheridentifikation* wird als ein Problem der Multiplen Detektion behandelt [2]. Die zu unterscheidenden Klassen, eine für jeden Sprecher, das von System erkannt werden soll, werden als sp_i $i = 1..M$ bezeichnet, mit M - Anzahl der von dem Sprechererkennungssystem erfassten Sprecher. Die Sprechererkennung basiert auf den aufgezeichneten Sprachsignalen der jeweiligen Sprecher. Das Sprachsignal wird segmentiert in die Signalrahmen $\mathbf{x} = [x(1)..x(K)]$ (z.B. für einen Signalrahmen von 20 ms Länge und eine Abtastfrequenz von 8 kHz beträgt K = 160). Die Segmentierung liefert die Sprachsignalrahmen $\mathbf{x}(1)..\mathbf{x}(N)$, wobei N von der Gesamtlänge des von dem Sprecher gesprochenes Satzes oder Schlüsselwortes abhängt. Die Entscheidung über den Sprecher wird aus den Wahrscheinlichkeiten oder Wahrscheinlichkeitsdichten (zusammen als Wahrscheinlichkeitsscores bezeichneten) getroffen, daß die Vektoren der Abtastwerte $\mathbf{x}(l)$ $l = 1..N$ der Klasse sp_i zugehören. Das statistisch optimale Entscheidungsschema wählt die Klasse sp_i mit dem höchsten Wahrscheinlichkeitswert bei gegebenen $\mathbf{x}(l)$, $l = 1..N$. D.h. der Vektor $\mathbf{x}(l)$ wird der Klasse sp_j zugeordnet, für die:

$$p(\mathbf{x}(1)...\mathbf{x}(N) \mid sp_j) > p(\mathbf{x}(1)...\mathbf{x}(N) \mid sp_i) \quad \text{für alle } j \neq i$$

Sprecherverifikation

Problem der *Sprecherverifikation* besteht darin, die vorgegebene Identität des Sprechers anhand seiner Stimme zu überprüfen. Dies entspricht der in Figur 2 (Problem der Sprecherverifikation) abgebildeten Situation.

Der Prozess der Sprecherverifikation verläuft auf ähnliche Weise wie der bei der Sprecheridentifikation, d.h. es wird

ebenfalls die Segmentierung des gesprochenen Satzes durchgeführt. Danach wird jedoch keine Klassifizierung der Stimme gemacht, sondern für die vorgegebene Sprecheridentität ein Wahrscheinlichkeitsscore berechnet und mit einer Schwelle verglichen. Die Identität des Sprechers wird also anhand seiner Stimme bestätigt, wenn:

$$p(\mathbf{x}(1)..\mathbf{x}(N) \mid sp_j) > schwelle$$

wobei sp_j der vorgegebenen Sprecheridentität entspricht. Die Schwelle muß entsprechend hoch gesetzt werden um die Situation zu vermeiden in der ein Sprecher mit einer anderer Identität als die vorgegebene zugelassen/autorisiert wird.

LPAS Kodierer

Die heute eingesetzten Sprachkodierverfahren basieren vorwiegend auf dem Analyse-durch-Synthese Verfahren mit einem LPC-Synthesefilter [2]. Die Sprachkodierung wird in diesen Verfahren durch Wiederholung der Kodierungs- und Dekodierungs-Operationen solange optimiert, bis der optimale Parametersatz für den gegebenen Sprachabschnitt gefunden wird.

Einer der am meisten verwendeten Typen des LPAS Kodierers ist der CELP Kodierer. Eine relativ neue Entwicklung ist der Harmonic Vector Excited Codec mit einer besonders für die beschriebene Aufgabe geeigneter Form der Anregungssignale. Synthesemodell eines CELP Kodierers ist in Figur 3 (Schema eines LPAS Kopierers) dargestellt. Das Synthesemodell definiert die Methode der Berechnung des synthetisierten Sprachsignals aus den quantisierten Parametern des Sprachsignals. Im allgemeinen besitzt jeder LPAS Kodierer besitzt Parametergruppen:

- Kurzzeitprädiktorparameter. Die Kurzzeitprädiktorparameter werden in der Regel mit Hilfe klassischer LPC Analyse berechnetet, wobei die Korrelations-Methode oder die Kovarianz-Methode der Linearen Prädiktion angewendet wird [3]. Für Signalrahmen der Länge von 20 bis 30 ms und eine Ab-

tastrate von 8kHz werden 8-10 LPC Koeffizienten verwendet. Die Kurzzeitprädiktorparameter können in verschiedenen Formen (z.B. die Reflexionkoeffizienten oder als Line Spectrum Frequencies LSF) auftreten, abhängig davon, welche Darstellung sich besser quantisieren läßt. Es hat sich gezeigt, daß die LSF Koeffizienten am besten zur Quantisierung geeignet sind und diese Form der Prädiktionskoeffizienten wird in der Regel verwendet. Die Kuzrzeitprädiktorparameter werden in einer open-loop Prozedur berechnet, d.h. ohne der in Figur 1 dargestellten gesamten Optimierung mit den anderen Parametern bezüglich des Synthesefehlers.

- Langzeitprädiktorparameter. Langzeitprädiktorparameter werden in einem Filter verwendet, der die Grundfrequenz des Sprachsignals synthetisiert. Es wird am meisten ein Langzeitprädiktor mit einem Filterkoeffizient und einem Parameter für die Grundperiode des Sprachsignals. Ein Langzeitprädiktor mit den Parametern $\mathbf{b} = [b, N]$ ist ein Teil der Figur 2. Die Langzeitprädiktorparameter werden ebenfalls in einer open-loop Prozedur berechnet ohne eine Gesamtoptimierung mit den anderen Parametern. In manchen Kodierern wird manchmal eine verfeinerte Suche nach den Langzeitprädiktorparametern in einer closed-loop Prozedur durchgeführt.

- Die Parameter der Anregung. In einem CELP Kodierer werden die 5-10ms Subrahmen des Restsignals in einer closed-loop Prozedur vektorquantisiert. Die gesendeten Parameter ermöglichen auf der Dekoderseite die Wiederherstellung der Signalformen aus dem gespeicherten Codebuch.

In einem HVXC Codecs wird der Ausgang aus dem LPC Analyse Filter in die Frequenzdomäne transformiert und die grundperiodennormalisierte Spektraleinhüllende vektorquantisiert.

Sprechererkennung mit den Parametern eines LPAS Kodierers

Die Parameter eines Sprachkodierers beschreiben ausführlich die möglichen Sprachsignale mit einer wesentlich reduzierten Anzahl der Parameter im Vergleich zur Darstellung des Sprachsignals als eine Sequenz der Abtastwerte.

Die Dekomposition des Sprachsignals in die erwähnten Parametergruppen kann auf verschiedene Weise zur Sprechererkennung verwendet werden. Die Methoden zur Berechnung der Parameter und Synthese des Sprachsignals implizieren die Methoden der Schätzung der Wahrscheinlichkeitsdichten (bzw. der Wahrscheinlichkeiten für die Parameter, die als diskrete Wahrscheinlichkeitsvariablen betrachtet werden). Die in einer closed-loop Prozedur bestimmt werden, sollen eigentlich als diskrete Wahrscheinlichkeitsvariablen betrachtet werden, weil es nicht möglich ist, für solche Parameter die Volumen der Parameterraumesregionen des Vektorquantisierers zu verbinden. Dies betrifft insbesondere die Anregungsparameter. Die Schätzung der Wahrscheinlichkeitsverteilungen für solche Parameter wird durch die Berechnung von relativen Häufigkeiten der Parameter/Codevektoren im Trainingssatz bestimmt.

Die in einer open-loop Prozedur im Kodierer berechnet werden, sind zuerst in einer nichtquantisierten Form verfügbar und dann erst quantisiert, wobei in der Regel die Vektorquantisierung verwendet wird. Für solche Parameter können die Wahrscheinlichkeitsdichten aus dem Trainingssatz geschätzt werden. Dieser Ansatz wird vor allem für die Kurzzeitprädiktorparameter angewendet.

Die Schätzung der Wahrscheinlichkeitsdichten basiert auf der Histogramm Methode [6]. Diese Methode benötigt die Kenntnisse der Volumen der mit den quantisierten Punkten verbundenen Regionen des Parameterraumes.

Eine Methode der Speicherung von Wahrscheinlichkeitsverteilungen ergibt sich gemäß Figur 5 (Sprecheridentifikation mit

den Parametern eines LPAS-Kopierers), wenn die möglichen Codevektoren für die Sprachsignalparameter einmal für die ganze Population gespeichert werden, was dem Fall entspricht, daß die Quantisierungsstufen/Codevektoren aus der Datenbank bestimmt, die die Aufzeichnungen von vielen Sprechern beinhaltet, einmal bestimmt werden. Die Wahrscheinlichkeitsverteilungen der Parameter für die Sprecher werden dann zusammen mit den Indizien der Codevektoren für die Parameter im System gespeichert. Sie eignet sich für große Systeme mit sehr vielen Anwendern (ATM, Zugangssysteme in Betrieben).

Eine andere Methode ergibt sich, wenn die Codevektoren für die Parameter für jeden Sprecher einzeln trainiert werden. Die Codevektoren werden dann zusammen mit den Werten der Wahrscheinlichkeitsdichten an den durch die Codevektoren bestimmten Punkten des Parameterraumes gespeichert. Ein Schema dieser Methode ist in Figur 6 (Sprecheridentifikation mit den Parametern eines LPAS Kopierers Wahrscheinlichkeitsdichten werden zusammen mit den Codevektoren für die Parameter gespeichert) gezeigt. Diese Methode ist für eine kleine Anzahl von Sprechern bestimmt (z.B. für eine mit der Stimme gesteuerte Tür in der Wohnung).

Trainingphase eines Sprechererkennungsystems

Die Wahrscheinlichkeitsdichteverteilungen für die Sprecherklassen werden aus dem Trainingsmaterial geschätzt. Für die textabhängige Sprechererkennung (Sprecheridentifikation/Sprecherverifikation) wird ein bestimmter Satz oder Schlüsselwort während der Trainingphase so lange wiederholt bis die Sprechererkennung sicher funktioniert.

Für die textunabhängige Sprecherverifikation muß ein phonetisch ausgewogenes Sprachmaterial aufgenommen werden. Auch in diesem Fall muß die Trainingphase solange wiederholt werden bis die Sprecheridentifikation/verifikation sicher funktioniert.

Das während der Trainingphase aufgenommene Material wird zum Training mehrmals jeweils phasenverschoben verwendet, um das Sprechererkennungssystem unabhängig von der Anfangsphase der aufgezeichneten Stimmen zu machen. Die zum Training verwendeten Daten wird als Trainingsatz TS_{sp_l} bezeichnet wobei sp_l den Sprecher symbolisiert.

Schätzung der Wahrscheinlichkeitsdichten

Um die erfindungsgemäße Methode zur Schätzung der Wahrscheinlichkeitsdichten der Parameter für die Sprecherklassen zu beschreiben, werden zuerst notwendige Definitionen eingeführt. Die eingeführte Abstraktion des Kodierungsprozesses hat den Vorteil, daß die Schätzung der Wahrscheinlichkeitsdichten auf einfache Weise beschrieben werden kann, ohne auf die sehr komplizierten Operationen im Sprachkodierer in Details einzugehen. Eine detaillierte Beschreibung der Parameterberechung kann in [4] und [5] gefunden werden.

Ein Sprachkodierer arbeitet in Auswerteinterwallen. Für jeden Signalrahmen werden in dem Sprachkodierer die im Abschnitt über LPAS Kodierer beschriebene Operationen durchgeführt, die die Parameter des Sprachsignals für den jeweiligen Rahmen liefern.

Berechnung eines nicht quantisierten Parametervektor $\mathbf{p}$ aus dem Signalrahmen $\mathbf{x}$ in einer open-loop Oprimierungsprozedur wird als $\mathbf{p} = K_p(\mathbf{x})$ geschrieben. Die Quantisierung des Parameters wird als: $\hat{\mathbf{p}} = Q_p(\mathbf{p})$ bezeichnet. Die Region im Parameterraum der Parameter $\mathbf{p}$, der im Kodierungsprozess auf den Codevektor $\hat{\mathbf{p}}$ abgebildet wird, wird als $S_{\hat{\mathbf{p}}} = \{\mathbf{p} : Q_p(\mathbf{p}) = \hat{\mathbf{p}}\}$ bezeichnet. Das Volumen von dieser Region wird als $V(S_{\hat{\mathbf{p}}})$ bezeichnet.

Der Satz möglicher Codevektoren für den Parameter $\mathbf{p}$ wird als $C_p = \{\hat{\mathbf{p}}_i; i = 1..N_p\}$ geschrieben mit N_p Anzahl von Codevektoren. Der Satz von Regionen, die mit den Codevektoren verbunden sind, wird als $R_p = \{S_i; i = 1..N_p\}$ bezeichnet. Die Zugehörigkeitsfunktion einer Region S_i wird als:

$$1_{S_i}(\mathbf{p}) = \begin{cases} 1 \text{ für } \mathbf{p} \in S_i \\ 0 \text{ für } \mathbf{p} \notin S_i \end{cases}$$

bezeichnet.

Die Häufigkeit des Vorkommens eines Parameters im Trainingssatz wird mit

$$f_{S_i} = \frac{\text{Anzanhl von Parameterwerten aus dem Training Satz } TS_{sp_i} \text{ die in den Region } S_i \text{ fallen}}{\text{Anzahl von Parameterwerten aus dem Training Satz } TS_{sp_i}}$$

berechnet.

Die geschätzte Wahrscheinlichkeitsdichteverteilung wird dann zu:

$$p(\mathbf{p} \mid sp_i) = \sum_{k=1}^{N_p} 1_{S_k}(\mathbf{p}) \frac{f_{S_i}}{V(S_l)}$$

Schätzung der Wahrscheinlichkeiten
Für die Parameter, die als eine diskrete Wahrscheinlichkeitsvariable betrachtet werden, d.h vor allem die Anregung aus dem Codebuch, die in einer closed-loop Prozedur optimiert wird und die Grundperiode des Sprachsignals, werden die Wahrscheinlichkeitsfunktionen (probability mass functions) geschätzt. Diese werden als die Häufigkeiten der gegebenen Parametercode im Trainingssatz für den jeweiligen Sprecher bestimmt.

Speichern der Wahrscheinlichkeitsverteilungen
Die Sprachparameter in einem Sprachkodierer werden nicht alle gleichzeitig sondern nacheinander berechnet. Es werden z.B. zuerst die Kurzzeitprädiktorparameter berechnet und dann für bereits bekannte Kurzzeitprädiktorparameter die restlichen Parameter bezüglich der Synthese oder des Prädiktionsfehlers optimiert. Dies ermöglicht effektives Speichern der Wahrscheinlichkeitsverteilungen als bedingte Wahrscheinlichkeiten der Codevektoren in einer Baumstruktur. Dies ist möglich dank folgender Abhängigkeit:

$$p(\mathbf{p}_K, \mathbf{p}_L, \mathbf{p}_A \mid sp_i) = p(\mathbf{p}_K \mid sp_i)p(\mathbf{p}_L \mid sp_i, \mathbf{p}_K)p(\mathbf{p}_A \mid sp_i, \mathbf{p}_K, \mathbf{p}_L)$$

$\mathbf{p}_K$ - Vektor von Kurzzeitparameter
$\mathbf{p}_L$ - Vektor von Langzeitparameter
$\mathbf{p}_A$ - Vektor von Anregungsparameter

Eine wesentliche Vereinfachung ergibt sich, wenn die Sprachparameter innerhalb eines Signalrahmens als statistisch unabhängig angenommen werden können. Die obige Formel wird dann zu:

$$p(\mathbf{p}_K, \mathbf{p}_L, \mathbf{p}_A \mid sp_i) = p(\mathbf{p}_K \mid sp_i)p(\mathbf{p}_L \mid sp_i)p(\mathbf{p}_A \mid sp)$$

Die Wahrscheinlichkeitsdichten müssen im System an sehr vielen Punkten im Parameterraum gespeichert werden. Die zum Speichern von Wahrscheinlichkeitsdichten verwendete Bitanzahl ist für die Komplexität des Gesamtsystems kritisch. Für die Wahrscheinlichkeitswerte wird deswegen ein Vektorquantisierer verwendet. Dies ermöglicht die Reduzierung der zum Speichern der Wahrscheinlichkeitsverteilungen verwendeten Bitanzahl.

Systemsicherheit
Um die Überlistung des Systems zu verhindern, wird gleichzeitig mit der Aufzeichnung der Stimme des Sprechers ein Rauschen ausgestrahlt, das dem System bekannt ist und aus dem das digitalisierte Sprachsignal subtrahiert wird.

5.
Die Erfindung kann für Anwendungen der Zutrittskontrolle, wie z.B. die mit der Stimme gesteuerte Tür, oder als Verifikation, beispielsweise für Bankzugangssysteme genutzt werden. Die Prozedur kann als ein Programmmodul auf einem Prozessor implementiert werden, der die Aufgabe der Sprechererkennung im System realisiert.

Ein Ausführungsbeispiel der Erfindung ist anhand der Figuren 7 und 8a bis 8m beschrieben.

[1] S. Furui, "Recent advances in speaker recognition", Pattern Recognition Letters, Tokyo Inst. of Technol., 1997

[2] P. Vary, U. Heute, W. Hess, *Digitale Sprachsignalverarbeitung*, B.G. Teubner Stuttgart, 1998

[3] K. Kroschel, *Statistische Nachrichtentheorie*, 3rd ed., Springer-Verlag, 1997

[4] W.B. Kleijn, K.K. Paliwal, *Speech Coding and Synthesis*, Elsevier, 1995

[5] ISO/IEC 14496-3, MPGA-3 HVXC Speech Coder description

[6] Prakasa Rao, Functional Estimation, Academic Press, 1982

Patentansprüche

1. Verfahren zum Erkennen von Sprechern anhand deren Stimmen mit folgenden Merkmalen:

(a) in einer Vorbereitungsphase,

(a1) werden von M Sprechern jeweils k textabhängige oder textunabhängige Referenzsprachäußerungen, die einen sprecherbezogenen Trainingssatz bilden, in erste Sprachsignalrahmen der Länge L segmentiert,

(a2) werden die ersten Sprachsignalrahmen einem auf linearer Prädiktion basierenden Analyse-durch-Synthese-Kodierer zugeführt,

(a3) wird in dem Analyse-durch-Synthese-Kodierer für jeden der M Sprecher und jeweils jeden ersten Sprachsignalrahmen ein erster Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter des Kodierers berechnet, wobei die Parameter dann ein sprecherbezogenes Trainingsmaterial bilden,

(a4) wird in dem Analyse-durch-Synthese-Kodierer für jeden der M Sprecher und jeweils jeden ersten Sprachsignalrahmen die Häufigkeit des jeweiligen Vorkommens des ersten Kurzzeitprädiktorparameters, Langzeitprädiktorparameters und/oder Anregungsparameters des Kodierers in dem sprecherbezogenen Trainingssatz bzw. die Wahrscheinlichkeitsdichten, mit der der erste Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter in dem sprecherbezogenen Trainingssatz enthalten ist, berechnet,

(a5) werden die berechneten Häufigkeiten bzw. Wahrscheinlichkeitsdichten sprecherbezogen als Sprecherdaten gespeichert,

(b) in einer simulierten Nutzungsphase der Trainingsphase,

(b1) wird eine textabhängige oder textunabhängige Simulationssprachäußerung eines m-ten Sprechers mit m=1..M in zweite Sprachsignalrahmen der Länge L segmentiert,

(b2) werden die zweiten Sprachsignalrahmen dem Analyse-durch-Synthese-Kodierer zugeführt,

(b3) wird in dem Analyse-durch-Synthese-Kodierer für den m-ten Sprecher und jeweils jeden zweiten Sprachsignalrahmen ein

zweiter Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter des Kodierers berechnet,
(b4) werden für jeden zweiten Sprachsignalrahmen aus dem berechneten zweiten Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter und den für den m-ten Sprecher in der Vorbereitungsphase gespeicherten Sprecherdaten erste Wahrscheinlichkeitstreffer berechnet, die angeben, mit welcher Wahrscheinlichkeit der zweite Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter mit dem ersten Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter übereinstimmt,
(b5) werden die ersten Wahrscheinlichkeitsscores aus allen zweiten Sprachsignalrahmen zusammengefaßt,
(b6) wird überprüft, ob die zusammengefaßten ersten Wahrscheinlichkeitsscores größer einer vorgegebenen ersten Schwelle sind, die Stimme des m-ten Sprechers bestätigt, wenn die zusammengefaßten ersten Wahrscheinlichkeitsscores größer als die vorgegebene erste Schwelle sind oder die Vorbereitungsphase solange für weitere i Referenzsprachäußerungen des m-ten Sprechers durchgeführt, bis die Stimme des m-ten Sprechers bestätigt wird, wenn die zusammengefaßten ersten Wahrscheinlichkeitsscores kleiner gleich oder kleiner der vorgegebenen ersten Schwelle sind,
(c) in einer Nutzungsphase
(c1) wird eine textabhängige oder textunabhängige Nutzsprachäußerung des m-ten Sprechers mit m=1.. M in dritte Sprachsignalrahmen der Länge L segmentiert,
(c2) werden die dritten Sprachsignalrahmen dem Analyse-durch-Synthese-Kodierer zugeführt,
(c3) wird in dem Analyse-durch-Synthese-Kodierer für den m-ten Sprecher und jeweils jeden dritten Sprachsignalrahmen ein dritter Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter des Kodierers berechnet,
(c4) werden für jeden dritten Sprachsignalrahmen aus dem berechneten dritten Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter und den für den m-

ten Sprecher in der Vorbereitungsphase gespeicherten Sprecherdaten zweite Wahrscheinlichkeitstreffer berechnet, die angeben, mit welcher Wahrscheinlichkeit der dritte Kurzzeitprädiktorparameter, Langzeitprädiktorparameter und/oder Anregungsparameter von dem m-ten Sprecher ausgesprochen wurde,

(c5) werden die zweiten Wahrscheinlichkeitstreffer aus allen dritten Sprachsignalrahmen zusammengefaßt,

(c6) wird überprüft, ob die zusammengefaßten zweiten Wahrscheinlichkeitsscores größer einer vorgegebenen zweiten Schwelle sind, die Stimme des m-ten Sprechers wird erkannt, wenn die zusammengefaßten zweiten Wahrscheinlichkeitstreffer größer der vorgegebenen zweiten Schwelle sind oder die Stimme des m-ten Sprechers wird nicht erkannt, wenn die zusammengefaßten zweiten Wahrscheinlichkeitsscores kleiner gleich oder kleiner der vorgegebenen zweiten Schwelle sind.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß
als ein parametrischer Kodierer, insbesondere ein "Harmonic Vector Excited Predictive"-Kodierer oder ein "Waveform Interpolating"-Kodierer verwendet wird.

3. Verfahren nach Anspruch 1, dadurch gekennzeichnet, daß
als Analyse-durch-Synthese-Kodierer ein auf linearer Prädiktion basierender Kodierer, insbesondere ein LPAS-Kodierer benutzt wird.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, daß
die Häufigkeiten bzw. Wahrscheinlichkeitsdichten mit einem Vektorquantisierer mit einer bestimmten, wesentlich reduzierten Bitanzahl quantisiert werden.

5. Verfahren nach einem der Ansprüche 1 bis 4, dadurch gekennzeichnet, daß

mit der Eingabe der Sprachäußerung des Sprechers in das Sprechererkennungssystem ein dem Sprechererkennungssystem bekanntes Rauschen mit eingegeben wird.

6. Verfahren nach einem der Ansprüche 1 bis 5, dadurch gekennzeichnet, daß
das miteingegebene Rauschen intern vor der Segmentierung von der Aufnahme der Sprecherstimme subtrahiert wird.

Zusammenfassung

Verfahren zum Erkennen von Sprechern anhand deren Stimmen

Die Erfindung betrifft ein Verfahren zur Sprechererkennung unter Anwendung von Parametern eines LPAS-Kopierers oder eines parametrischen Kopierers zur Modellierung der Wahrscheinlichkeitsverteilung für die Sprecherklassen.

FIG 1

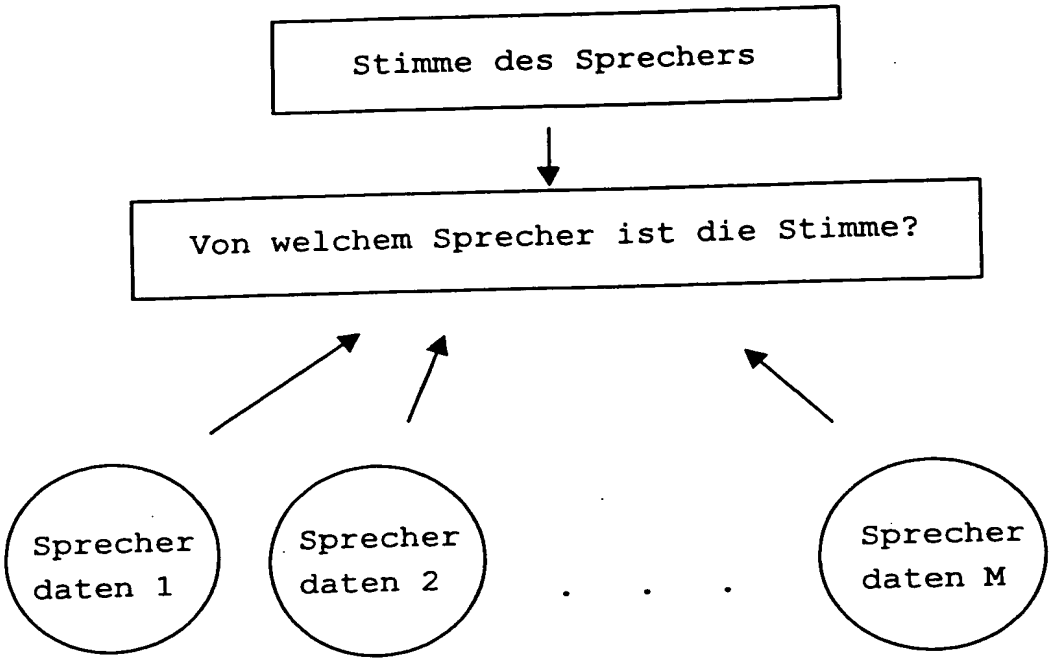

Stimme des Sprechers
Von welchem Sprecher ist die Stimme?
Sprecher daten 1
Sprecher daten 2
. . .
Sprecher daten M

FIG 2

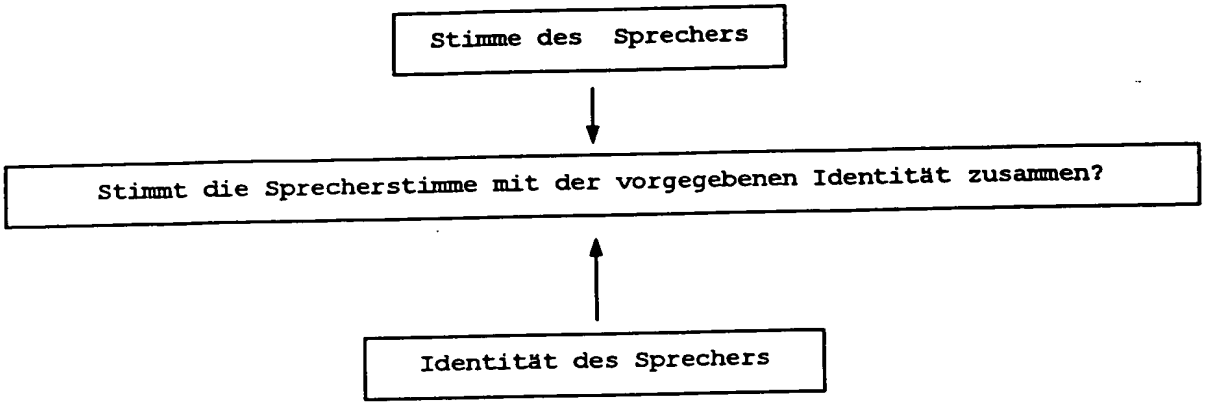

Stimme des Sprechers
Stimmt die Sprecherstimme mit der vorgegebenen Identität zusammen?
Identität des Sprechers

FIG 3

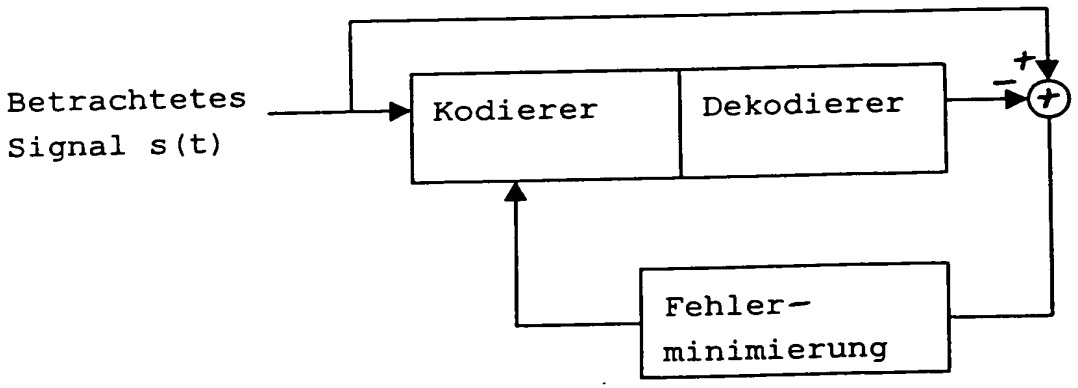

Betrachtetes
Signal s(t)
Kodierer
Dekodierer
+
−
+
Fehler-
minimierung

FIG 4

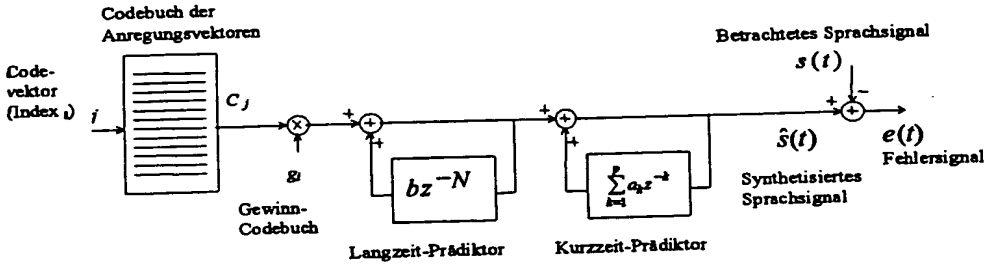

Codebuch der Anregungsvektoren
Code-vektor (Index j)
j
c_j
g
Gewinn-Codebuch
bz^{-N}
Langzeit-Prädiktor
$\sum_{k=1}^{p} a_k z^{-k}$
Kurzzeit-Prädiktor
Betrachtetes Sprachsignal
$s(t)$
$\hat{s}(t)$
Synthetisiertes Sprachsignal
$e(t)$
Fehlersignal

FIG 5

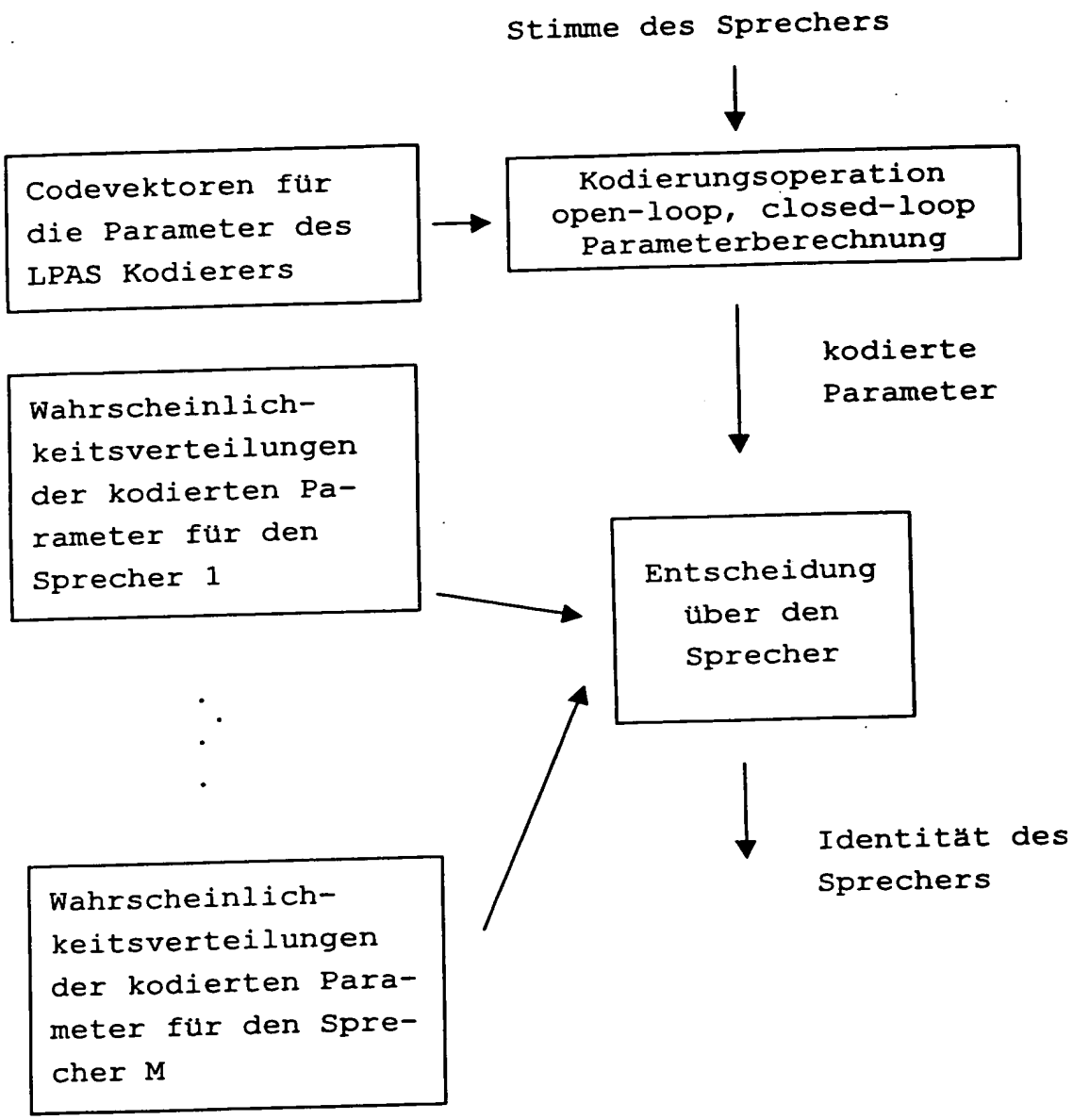

Stimme des Sprechers
Codevektoren für die Parameter des LPAS Kodierers
Kodierungsoperation open-loop, closed-loop Parameterberechnung
kodierte Parameter
Wahrscheinlich-keitsverteilungen der kodierten Pa-rameter für den Sprecher 1
Entscheidung über den Sprecher
Identität des Sprechers
Wahrscheinlich-keitsverteilungen der kodierten Para-meter für den Spre-cher M

FIG 6

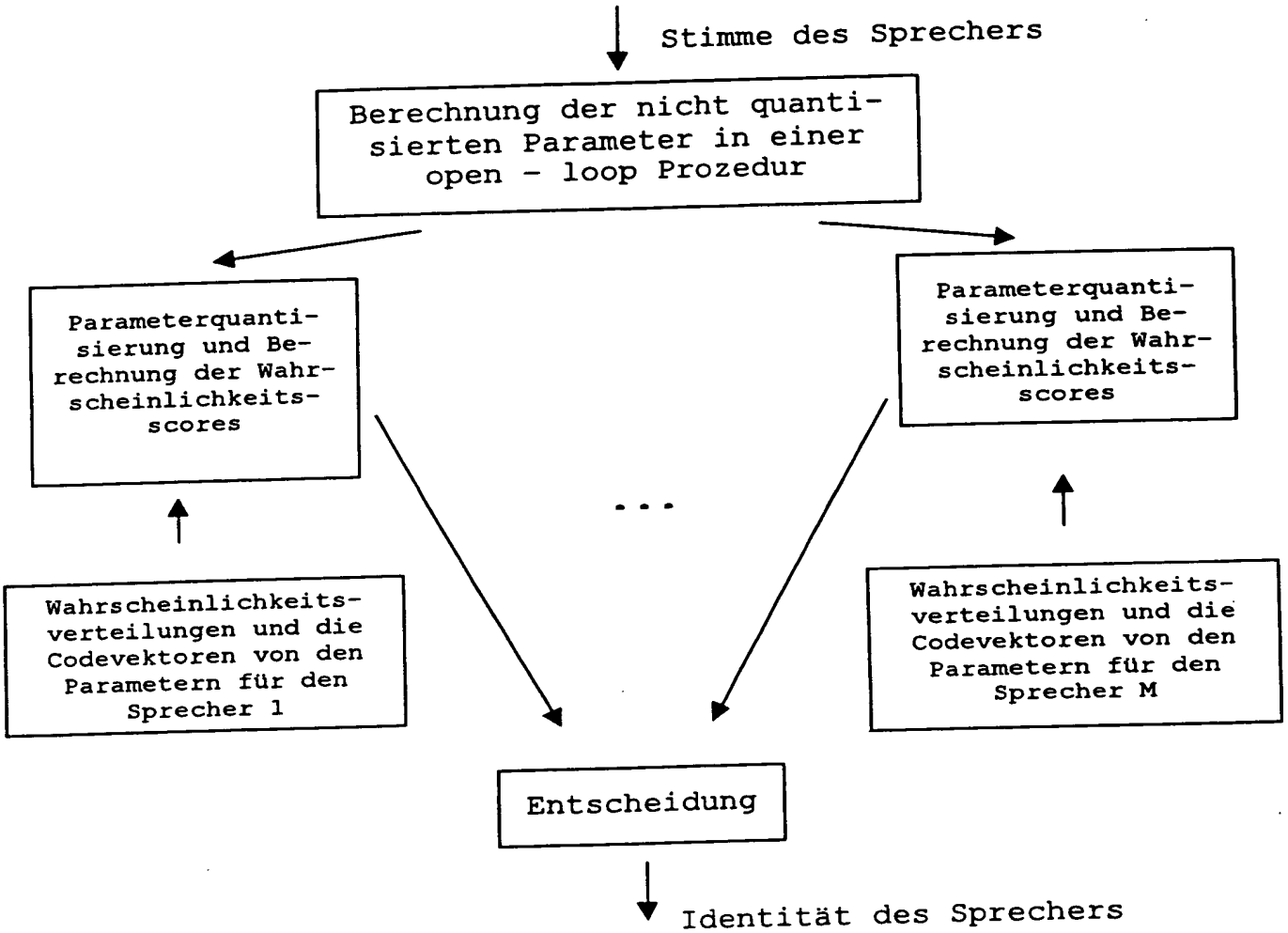

Stimme des Sprechers
Berechnung der nicht quantisierten Parameter in einer open - loop Prozedur
Parameterquantisierung und Berechnung der Wahrscheinlichkeitsscores
Parameterquantisierung und Berechnung der Wahrscheinlichkeitsscores
...
Wahrscheinlichkeitsverteilungen und die Codevektoren von den Parametern für den Sprecher 1
Wahrscheinlichkeitsverteilungen und die Codevektoren von den Parametern für den Sprecher M
Entscheidung
Identität des Sprechers

FIG 7

Sprecherverifikation mit Verwendung von den Parametern eines LPAS Kodierers

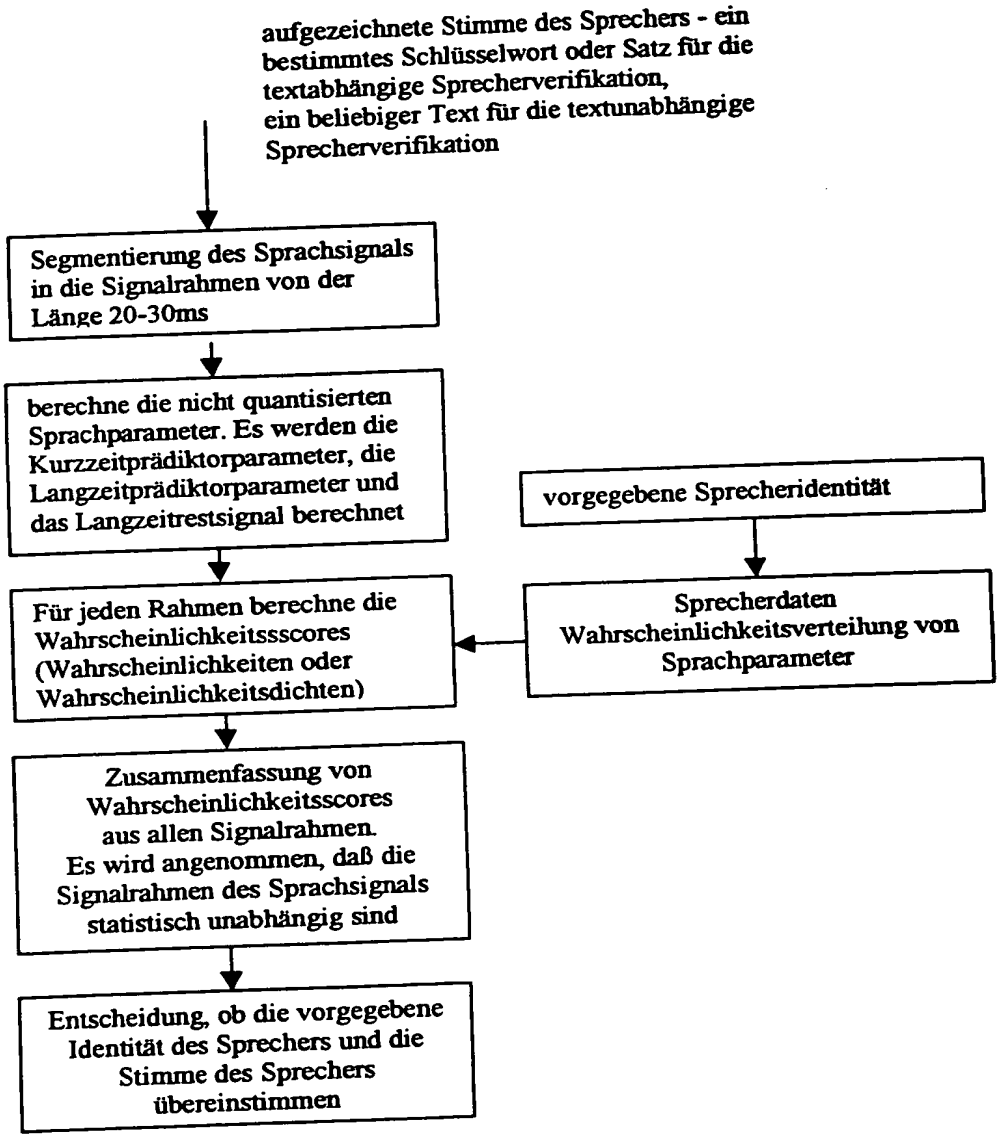
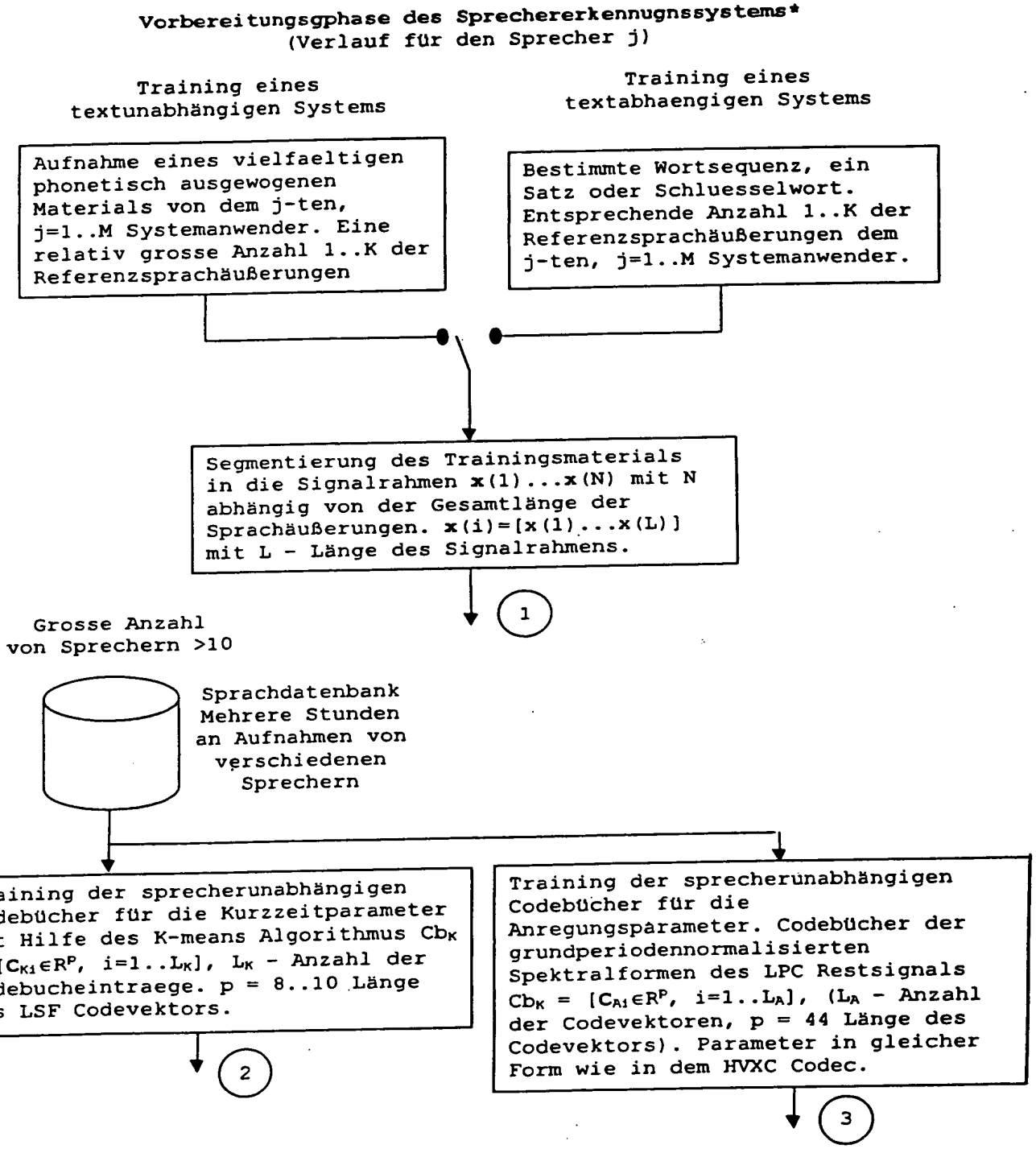
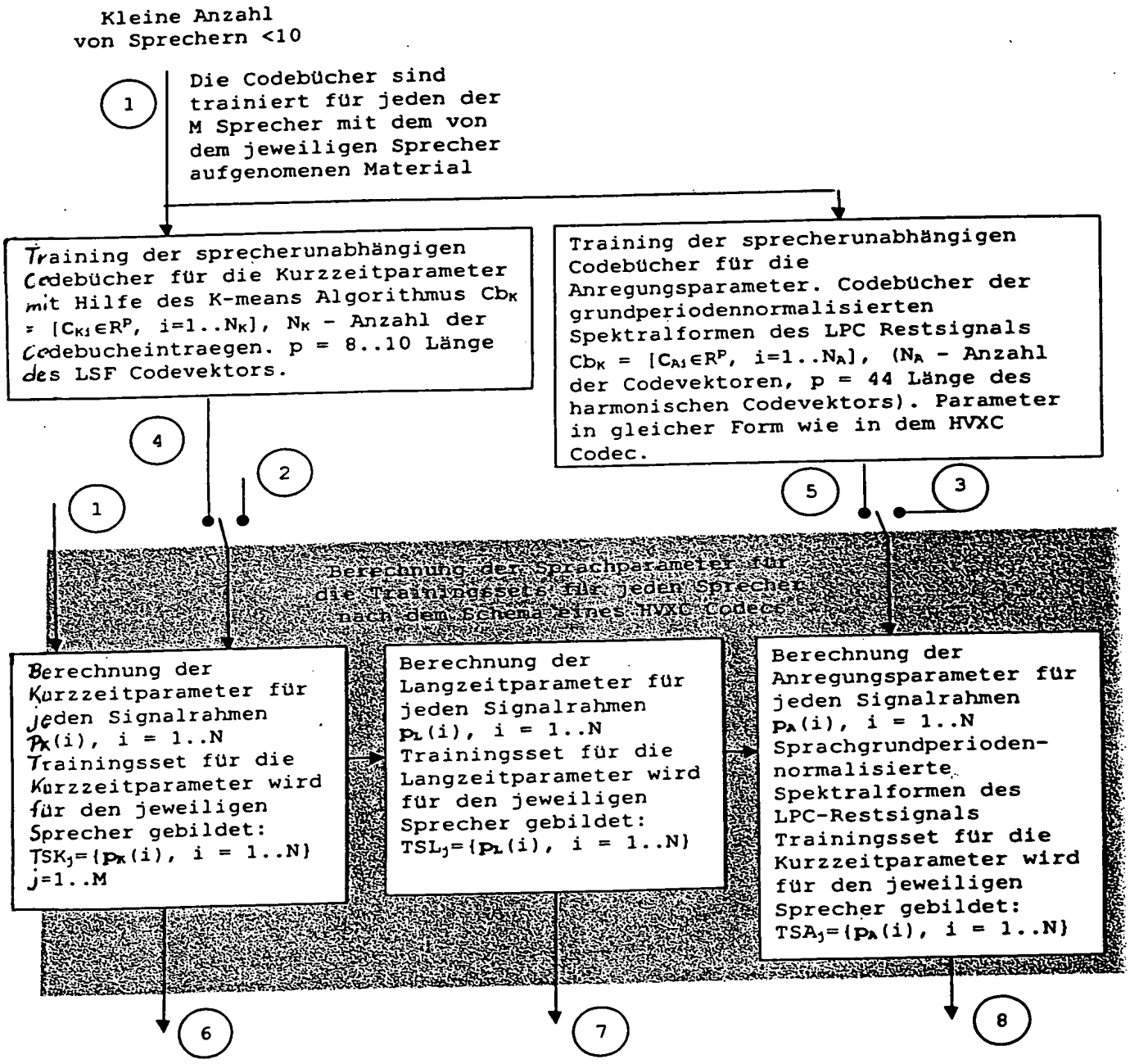


FIG 8a



* Der im folgenden definierte Prozess wird für jeden neuen Nutzer des Sprechererkennungssystems durchgeführt. Das Ziel der Vorbereitungsphase ist die Erstellung der Sprecherdaten für jeden der M Sprecher.

FIG 8b



* ISO/IEC 14496-3 Information Technology - Very Low Bitrate Audio-Visual Coding

FIG 8c

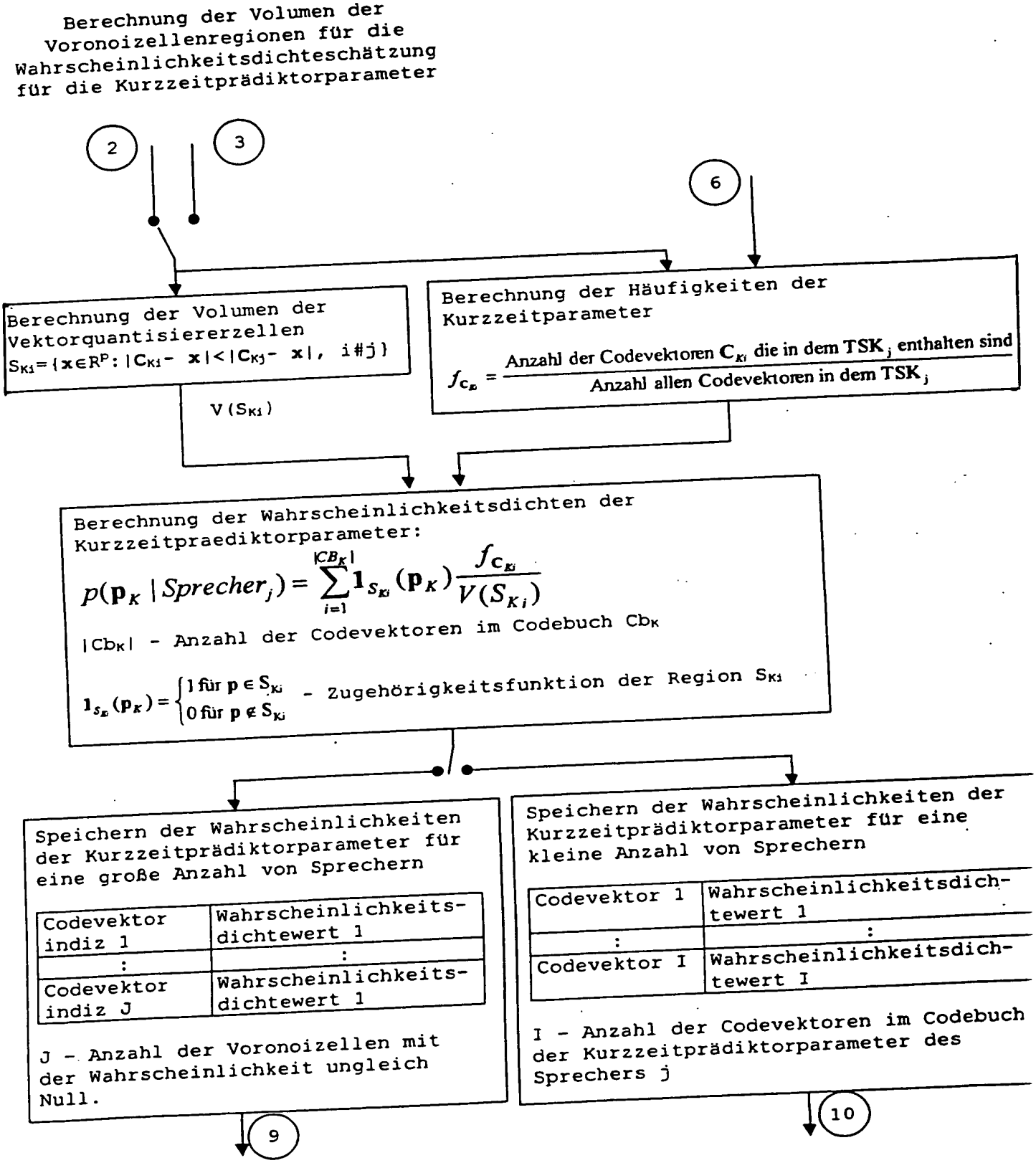

Berechnung der Volumen der Voronoizellenregionen für die Wahrscheinlichkeitsdichteschätzung für die Kurzzeitprädiktorparameter
2
3
6
Berechnung der Volumen der Vektorquantisiererzellen
$S_{Ki}=\{x\in R^P: |C_{Ki}- x|<|C_{Kj}- x|, i\#j\}$
$V(S_{Ki})$
Berechnung der Häufigkeiten der Kurzzeitparameter
$f_{C_{Ki}} = \frac{\text{Anzahl der Codevektoren } C_{Ki} \text{ die in dem TSK}_j \text{ enthalten sind}}{\text{Anzahl allen Codevektoren in dem TSK}_j}$
Berechnung der Wahrscheinlichkeitsdichten der Kurzzeitpraediktorparameter:
$p(\mathbf{p}_K \mid Sprecher_j) = \sum_{i=1}^{|CB_K|} \mathbf{1}_{S_{Ki}}(\mathbf{p}_K) \frac{f_{C_{Ki}}}{V(S_{Ki})}$
$|Cb_K|$ - Anzahl der Codevektoren im Codebuch Cb_K
$\mathbf{1}_{S_{Ki}}(\mathbf{p}_K) = \begin{cases} 1 \text{ für } \mathbf{p} \in S_{Ki} \\ 0 \text{ für } \mathbf{p} \notin S_{Ki} \end{cases}$ - Zugehörigkeitsfunktion der Region S_{Ki}
Speichern der Wahrscheinlichkeiten der Kurzzeitprädiktorparameter für eine große Anzahl von Sprechern
Codevektor indiz 1 | Wahrscheinlichkeits-dichtewert 1
:
Codevektor indiz J | Wahrscheinlichkeits-dichtewert 1
J - Anzahl der Voronoizellen mit der Wahrscheinlichkeit ungleich Null.
Speichern der Wahrscheinlichkeiten der Kurzzeitprädiktorparameter für eine kleine Anzahl von Sprechern
Codevektor 1 | Wahrscheinlichkeitsdich-tewert 1
:
Codevektor I | Wahrscheinlichkeitsdich-tewert I
I - Anzahl der Codevektoren im Codebuch der Kurzzeitprädiktorparameter des Sprechers j
9
10

FIG 8d

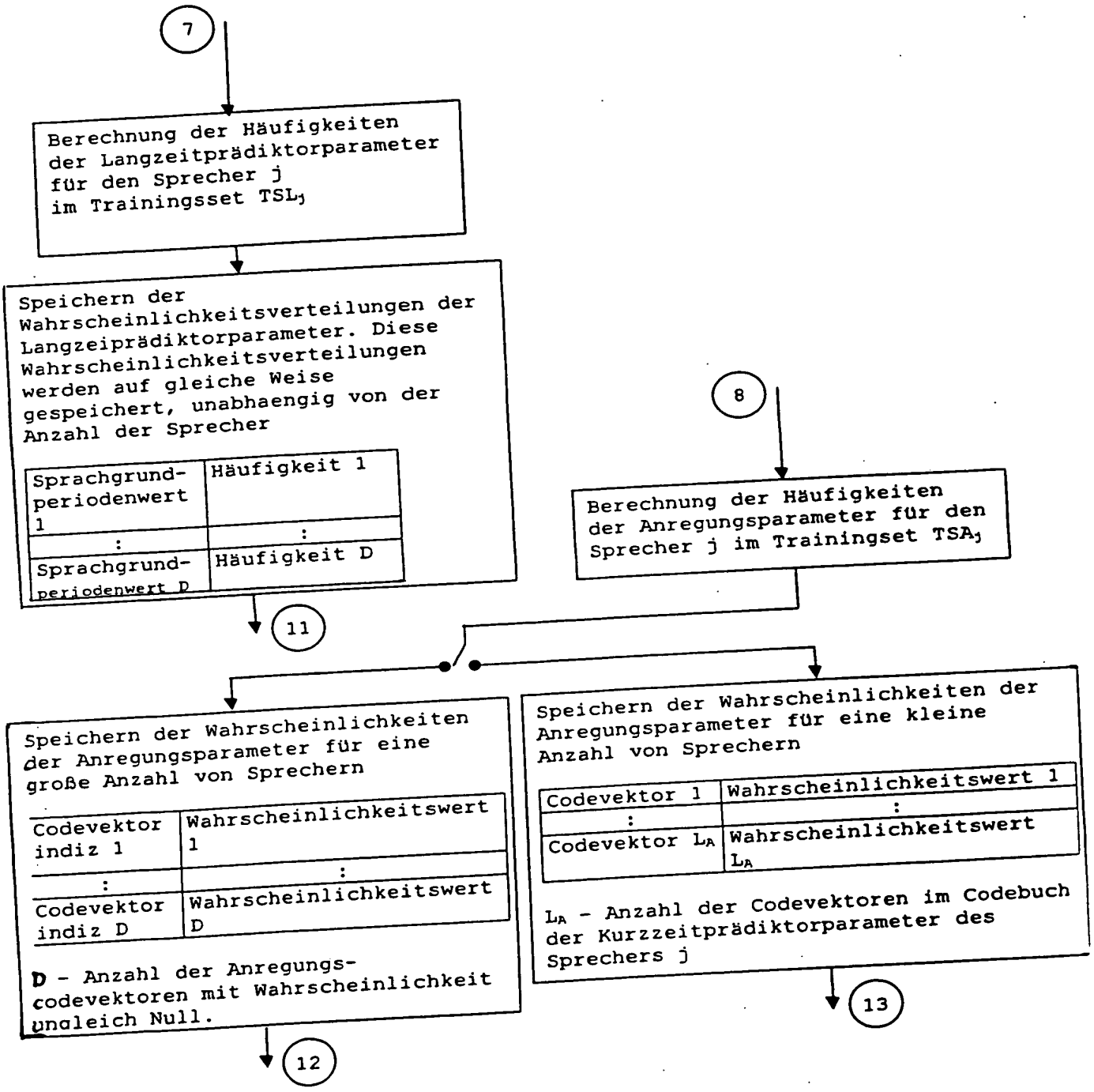

7
Berechnung der Häufigkeiten der Langzeitprädiktorparameter für den Sprecher j im Trainingsset TSL_j
Speichern der Wahrscheinlichkeitsverteilungen der Langzeiprädiktorparameter. Diese Wahrscheinlichkeitsverteilungen werden auf gleiche Weise gespeichert, unabhaengig von der Anzahl der Sprecher
Sprachgrund-periodenwert 1 | Häufigkeit 1
: | :
Sprachgrund-periodenwert D | Häufigkeit D
11
8
Berechnung der Häufigkeiten der Anregungsparameter für den Sprecher j im Trainingset TSA_j
Speichern der Wahrscheinlichkeiten der Anregungsparameter für eine große Anzahl von Sprechern
Codevektor indiz 1 | Wahrscheinlichkeitswert 1
: | :
Codevektor indiz D | Wahrscheinlichkeitswert D
D - Anzahl der Anregungs-codevektoren mit Wahrscheinlichkeit ungleich Null.
12
Speichern der Wahrscheinlichkeiten der Anregungsparameter für eine kleine Anzahl von Sprechern
Codevektor 1 | Wahrscheinlichkeitswert 1
: | :
Codevektor L_A | Wahrscheinlichkeitswert L_A
L_A - Anzahl der Codevektoren im Codebuch der Kurzzeitprädiktorparameter des Sprechers j
13

FIG 8

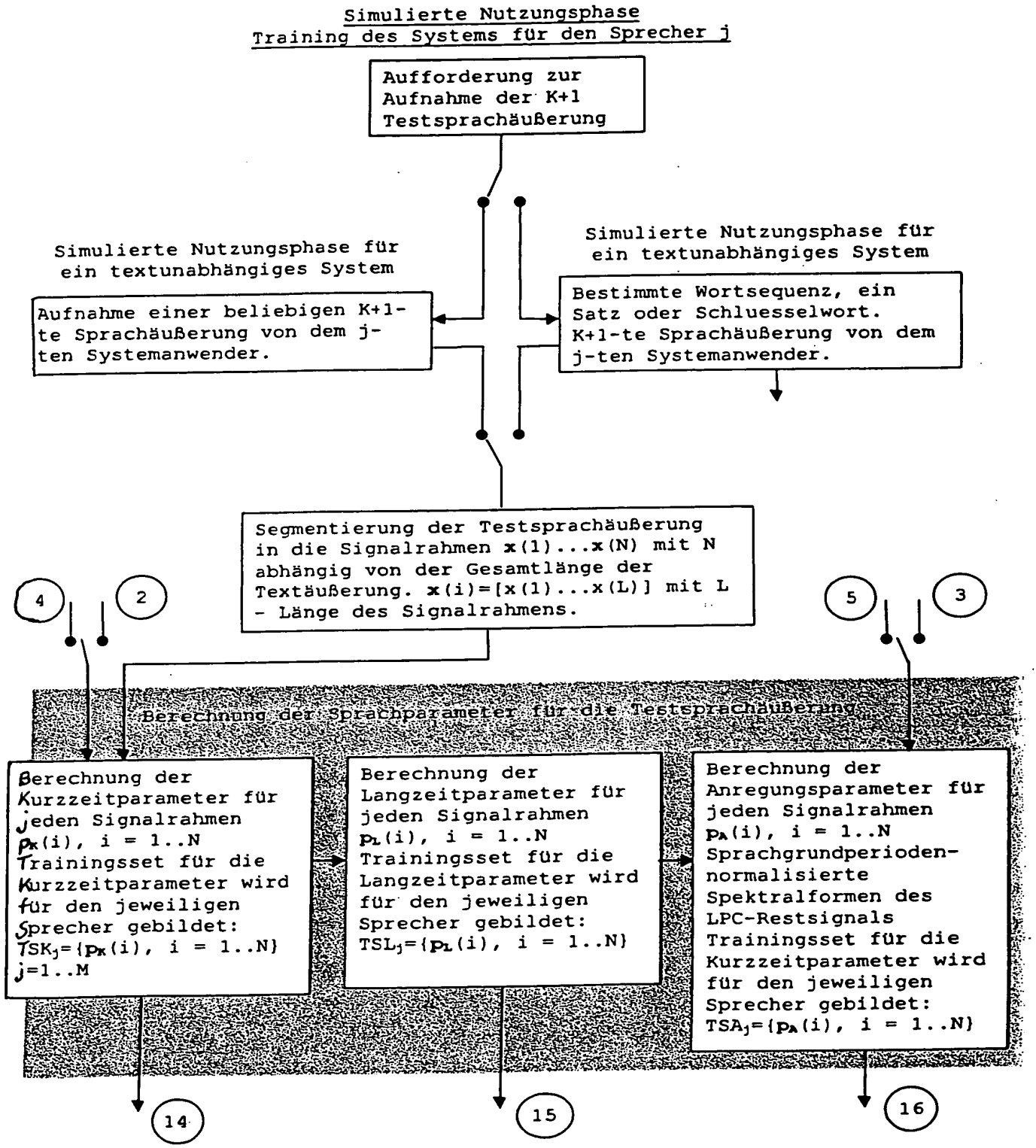

Simulierte Nutzungsphase
Training des Systems für den Sprecher j
Aufforderung zur Aufnahme der K+1 Testsprachäußerung
Simulierte Nutzungsphase für ein textunabhängiges System
Aufnahme einer beliebigen K+1-te Sprachäußerung von dem j-ten Systemanwender.
Simulierte Nutzungsphase für ein textunabhängiges System
Bestimmte Wortsequenz, ein Satz oder Schluesselwort. K+1-te Sprachäußerung von dem j-ten Systemanwender.
Segmentierung der Testsprachäußerung in die Signalrahmen x(1)...x(N) mit N abhängig von der Gesamtlänge der Textäußerung. x(i)=[x(1)...x(L)] mit L - Länge des Signalrahmens.
4
2
5
3
Berechnung der Sprachparameter für die Testsprachäußerung
Berechnung der Kurzzeitparameter für jeden Signalrahmen $p_K(i)$, i = 1..N Trainingsset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: $TSK_j=\{p_K(i), i = 1..N\}$ j=1..M
Berechnung der Langzeitparameter für jeden Signalrahmen $p_L(i)$, i = 1..N Trainingsset für die Langzeitparameter wird für den jeweiligen Sprecher gebildet: $TSL_j=\{p_L(i), i = 1..N\}$
Berechnung der Anregungsparameter für jeden Signalrahmen $p_A(i)$, i = 1..N Sprachgrundperioden-normalisierte Spektralformen des LPC-Restsignals Trainingsset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: $TSA_j=\{p_A(i), i = 1..N\}$
14
15
16

FIG 8f

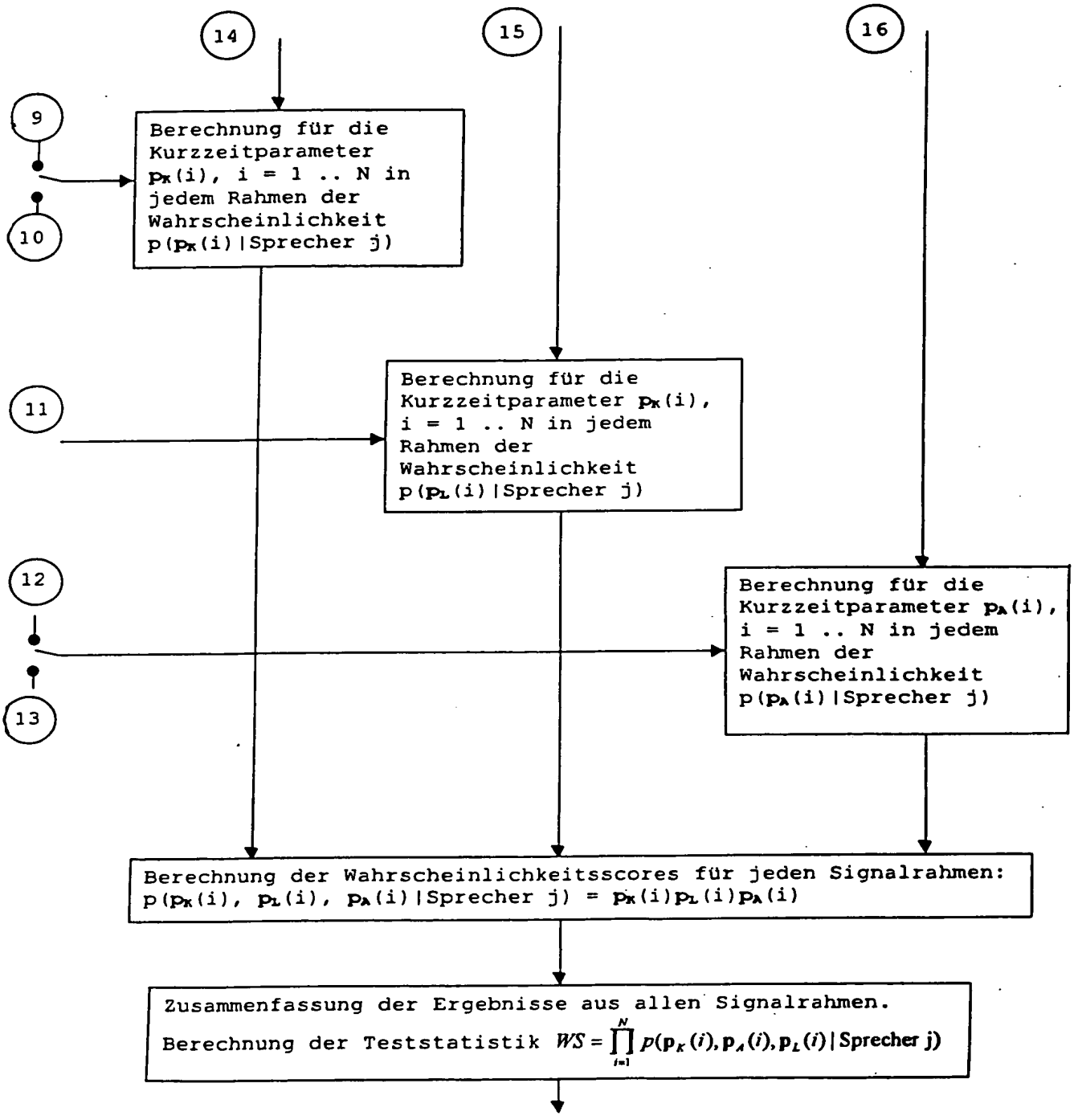
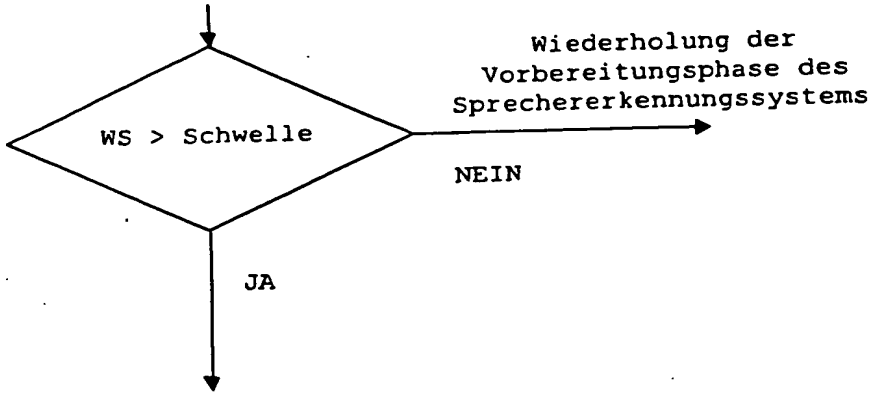

14
15
16
9
10
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_K(i)|\text{Sprecher j})$
11
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_L(i)|\text{Sprecher j})$
12
13
Berechnung für die Kurzzeitparameter $p_A(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_A(i)|\text{Sprecher j})$
Berechnung der Wahrscheinlichkeitsscores für jeden Signalrahmen: $p(p_K(i), p_L(i), p_A(i)|\text{Sprecher j}) = p_K(i)p_L(i)p_A(i)$
Zusammenfassung der Ergebnisse aus allen Signalrahmen.
Berechnung der Teststatistik $WS = \prod_{i=1}^{N} p(\mathbf{p}_K(i), \mathbf{p}_A(i), \mathbf{p}_L(i)|\text{Sprecher j})$

FIG 8g



kein zusätzliches Training der Wahrscheinlichkeitsverteilungen nötig. Die Wahrscheinlichkeitsverteilungen werden im System gespeichert und sind fertig für die Nutzungphase.

FIG 8h

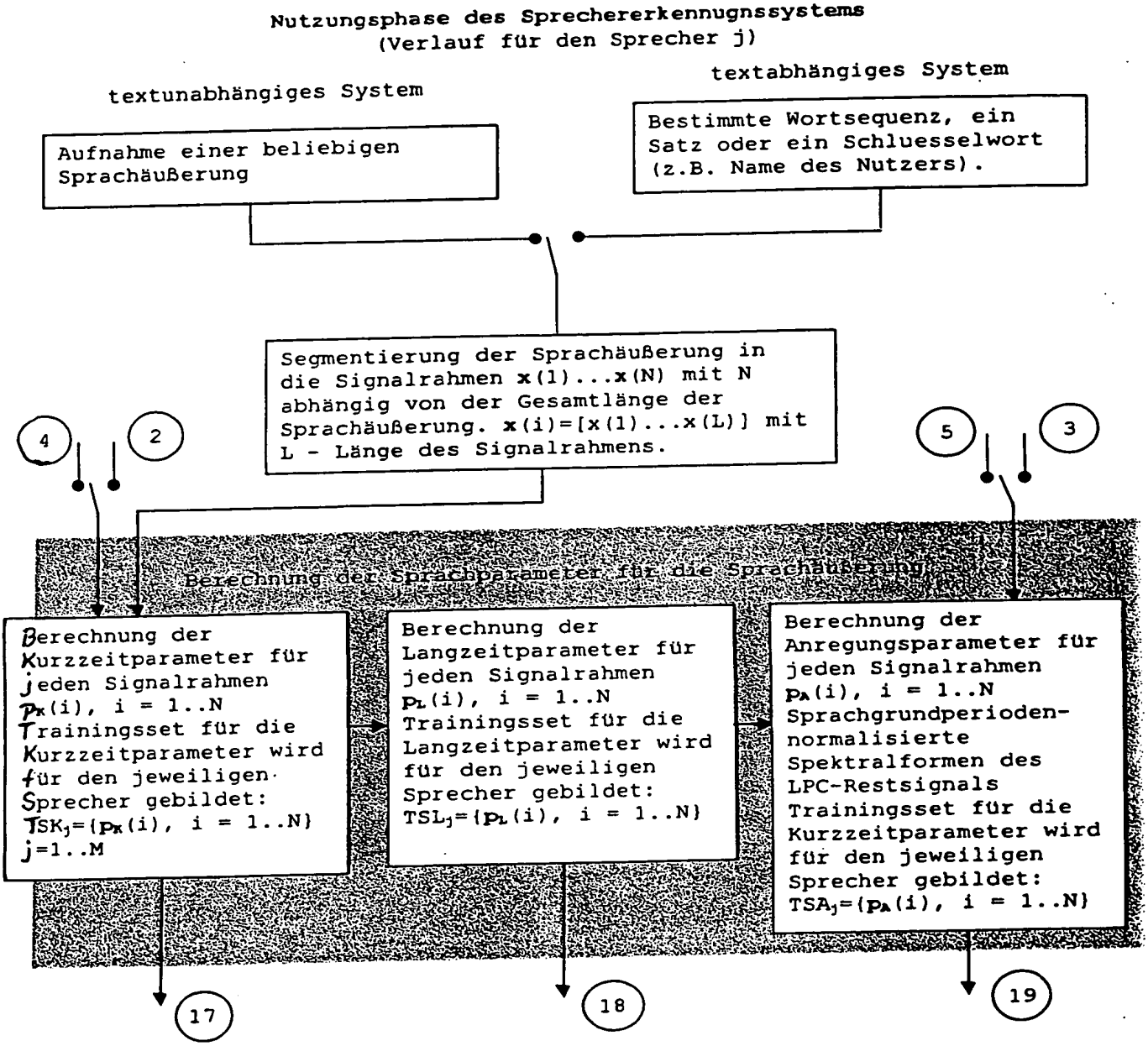

Nutzungsphase des Sprechererkennugnssystems
(Verlauf für den Sprecher j)
textunabhängiges System
textabhängiges System
Aufnahme einer beliebigen Sprachäußerung
Bestimmte Wortsequenz, ein Satz oder ein Schluesselwort (z.B. Name des Nutzers).
Segmentierung der Sprachäußerung in die Signalrahmen x(1)...x(N) mit N abhängig von der Gesamtlänge der Sprachäußerung. x(i)=[x(1)...x(L)] mit L - Länge des Signalrahmens.
4
2
5
3
Berechnung der Sprachparameter für die Sprachäußerung
Berechnung der Kurzzeitparameter für jeden Signalrahmen $p_K(i)$, i = 1..N Trainingsset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: $TSK_j=\{p_K(i), i = 1..N\}$ j=1..M
Berechnung der Langzeitparameter für jeden Signalrahmen $p_L(i)$, i = 1..N Trainingsset für die Langzeitparameter wird für den jeweiligen Sprecher gebildet: $TSL_j=\{p_L(i), i = 1..N\}$
Berechnung der Anregungsparameter für jeden Signalrahmen $p_A(i)$, i = 1..N Sprachgrundperioden-normalisierte Spektralformen des LPC-Restsignals Trainingsset für die Kurzzeitparameter wird für den jeweiligen Sprecher gebildet: $TSA_j=\{p_A(i), i = 1..N\}$
17
18
19

FIG 8i

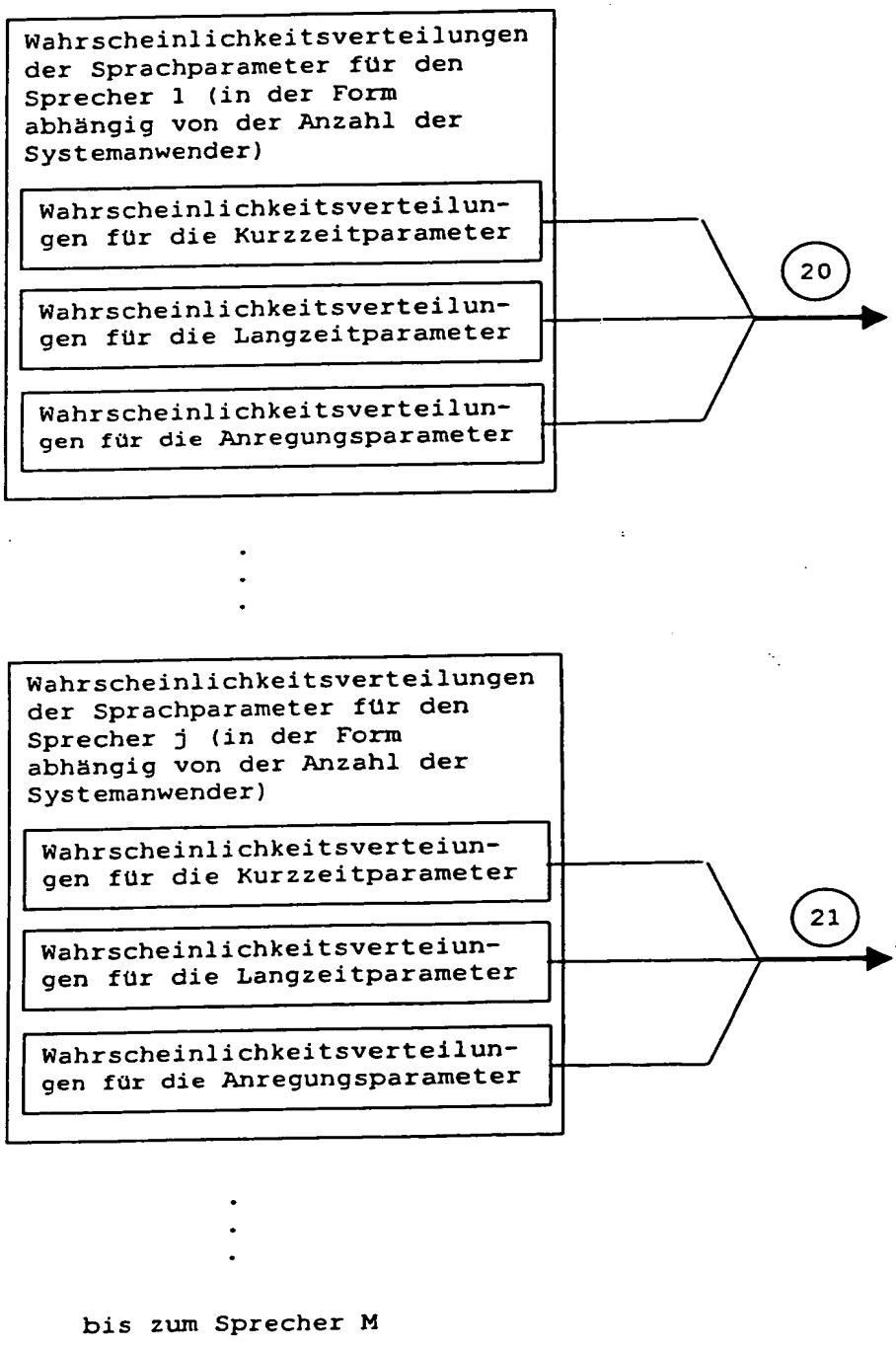

Wahrscheinlichkeitsverteilungen der Sprachparameter für den Sprecher 1 (in der Form abhängig von der Anzahl der Systemanwender)
Wahrscheinlichkeitsverteilungen für die Kurzzeitparameter
Wahrscheinlichkeitsverteilungen für die Langzeitparameter
Wahrscheinlichkeitsverteilungen für die Anregungsparameter
20
Wahrscheinlichkeitsverteilungen der Sprachparameter für den Sprecher j (in der Form abhängig von der Anzahl der Systemanwender)
Wahrscheinlichkeitsverteiungen für die Kurzzeitparameter
Wahrscheinlichkeitsverteiungen für die Langzeitparameter
Wahrscheinlichkeitsverteilungen für die Anregungsparameter
21
bis zum Sprecher M

FIG 8j

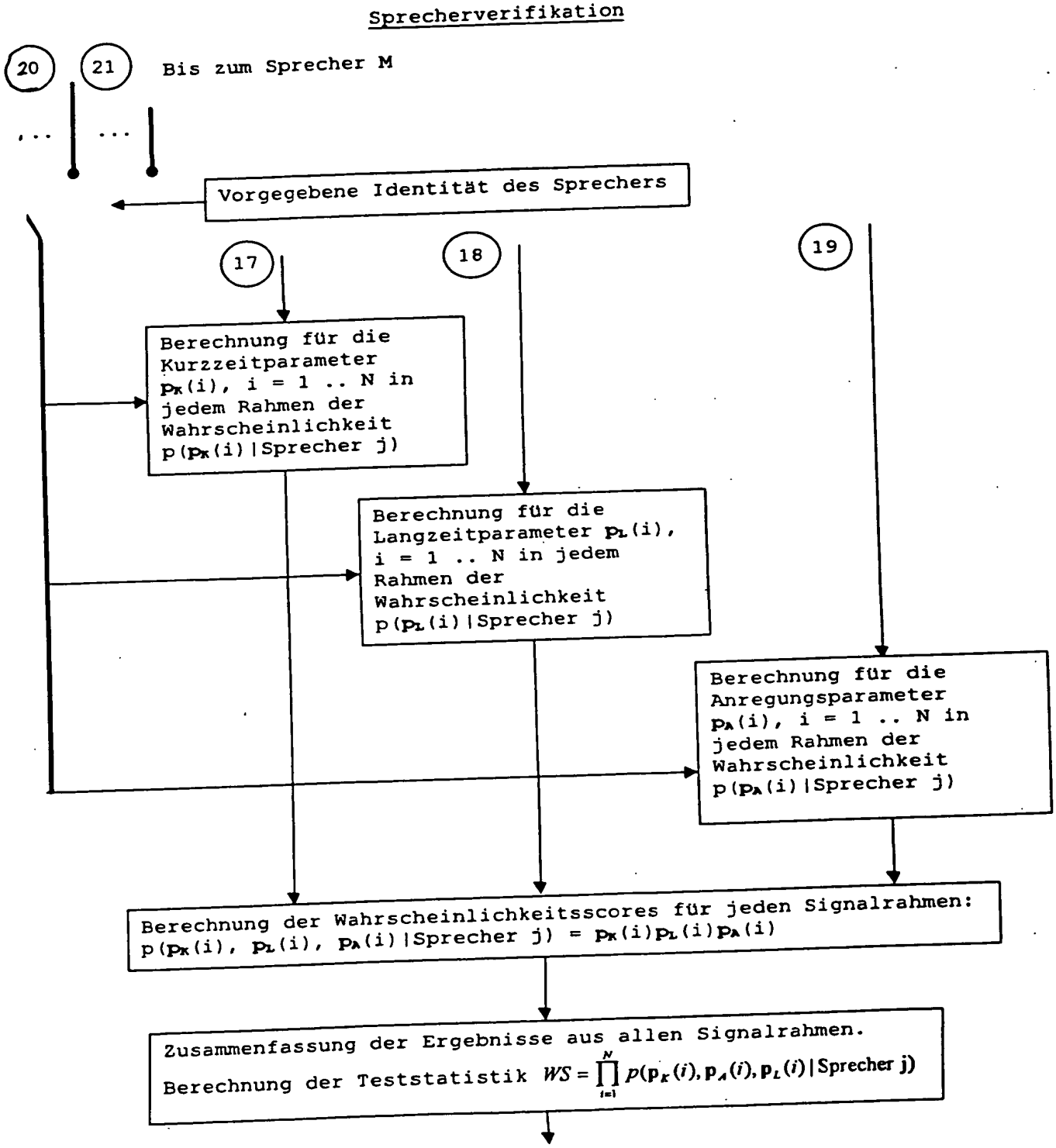

Sprecherverifikation
20
21
Bis zum Sprecher M
Vorgegebene Identität des Sprechers
17
18
19
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_K(i)|\text{Sprecher j})$
Berechnung für die Langzeitparameter $p_L(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_L(i)|\text{Sprecher j})$
Berechnung für die Anregungsparameter $p_A(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeit $p(p_A(i)|\text{Sprecher j})$
Berechnung der Wahrscheinlichkeitsscores für jeden Signalrahmen: $p(p_K(i), p_L(i), p_A(i)|\text{Sprecher j}) = p_K(i)p_L(i)p_A(i)$
Zusammenfassung der Ergebnisse aus allen Signalrahmen.
Berechnung der Teststatistik $WS = \prod_{i=1}^{N} p(p_K(i), p_A(i), p_L(i)|\text{Sprecher j})$

FIG 8k

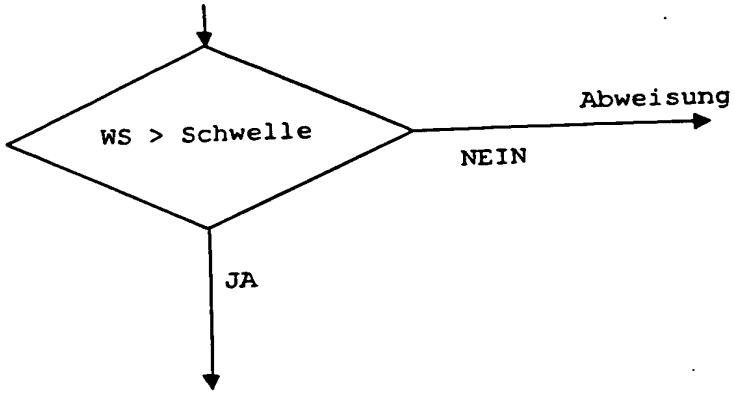

WS > Schwelle
Abweisung
NEIN
JA
Bestätigung der Sprecheridentität

FIG 81

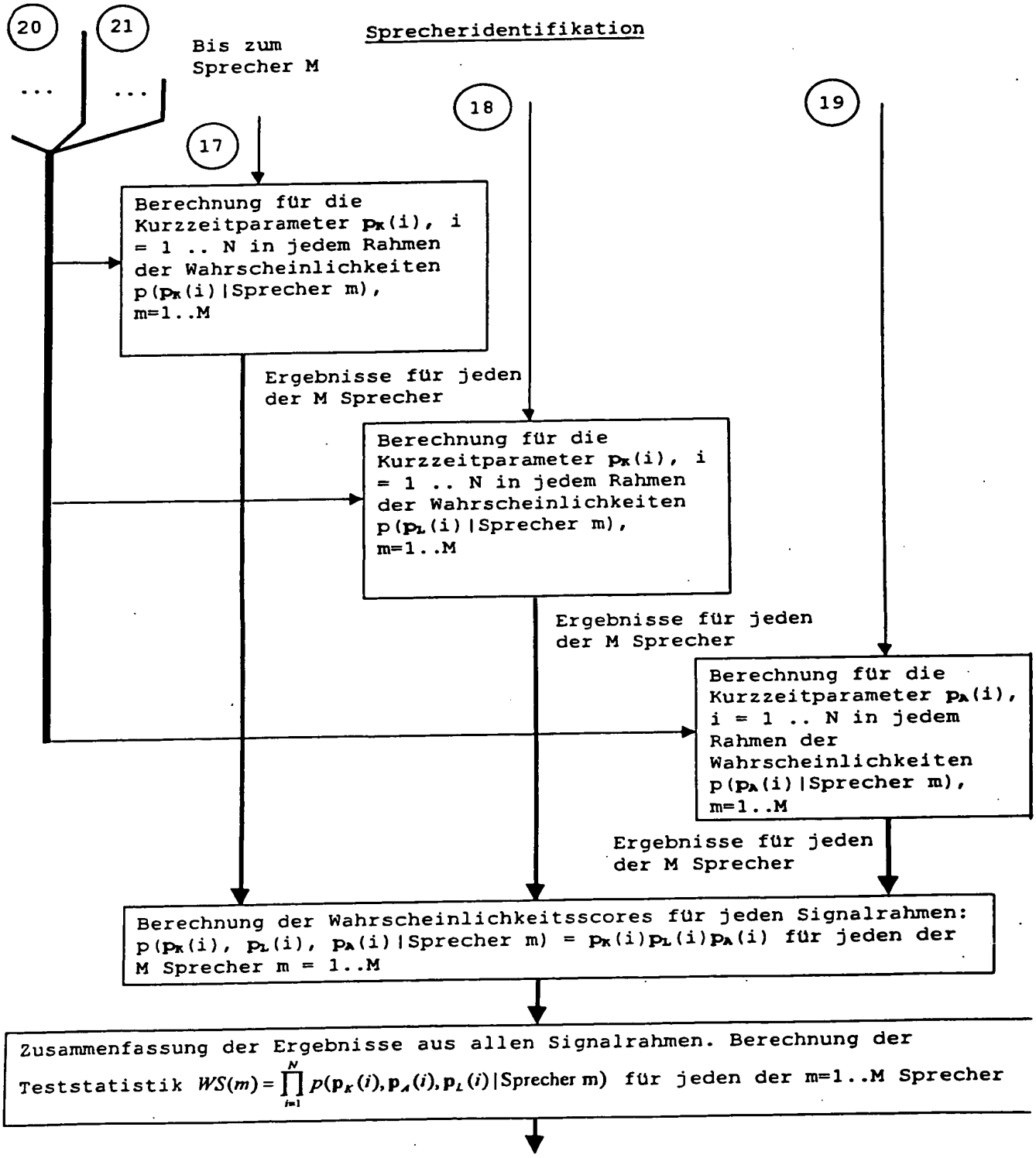
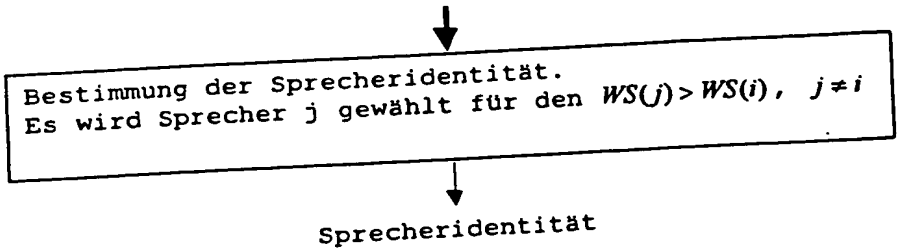

Sprecheridentifikation
20
21
Bis zum Sprecher M
17
18
19
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeiten $p(p_K(i)|\text{Sprecher m})$, m=1..M
Ergebnisse für jeden der M Sprecher
Berechnung für die Kurzzeitparameter $p_K(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeiten $p(p_L(i)|\text{Sprecher m})$, m=1..M
Ergebnisse für jeden der M Sprecher
Berechnung für die Kurzzeitparameter $p_A(i)$, i = 1 .. N in jedem Rahmen der Wahrscheinlichkeiten $p(p_A(i)|\text{Sprecher m})$, m=1..M
Ergebnisse für jeden der M Sprecher
Berechnung der Wahrscheinlichkeitsscores für jeden Signalrahmen: $p(p_K(i), p_L(i), p_A(i)|\text{Sprecher m}) = p_K(i)p_L(i)p_A(i)$ für jeden der M Sprecher m = 1..M
Zusammenfassung der Ergebnisse aus allen Signalrahmen. Berechnung der Teststatistik $WS(m) = \prod_{i=1}^{N} p(\mathbf{p}_K(i), \mathbf{p}_A(i), \mathbf{p}_L(i)|\text{Sprecher m})$ für jeden der m=1..M Sprecher

FIG 8m


Bestimmung der Sprecheridentität.
Es wird Sprecher j gewählt für den $WS(j) > WS(i)$, $j \neq i$
Sprecheridentität